BIOPHYSICS

VASANTHA PATTABHI
N. GAUTHAM

VISIT…

Biophysics

This Page Intentionally Left Blank

Biophysics

Vasantha Pattabhi
N. Gautham

Department of Crystallography & Biophysics
University of Madras, Guindy Campus
Chennai, India

KLUWER ACADEMIC PUBLISHERS
NEW YORK, BOSTON, DORDRECHT, LONDON, MOSCOW

Narosa Publishing House
DELHI CHENNAI MUMBAI KOLKATA

eBook ISBN: 0-306-47520-0
Print ISBN: 1-4020-0218-1

©2002 Kluwer Academic Publishers
New York, Boston, Dordrecht, London, Moscow

Print ©2002 Kluwer Academic Publishers
Dordrecht

All rights reserved

No part of this eBook may be reproduced or transmitted in any form or by any means, electronic, mechanical, recording, or otherwise, without written consent from the Publisher

Created in the United States of America

Visit Kluwer Online at: http://kluweronline.com
and Kluwer's eBookstore at: http://ebooks.kluweronline.com

This book is dedicated to

Professor G.N. Ramachandran, F.R.S.

who continues to inspire generations of biophysicists

எப்பொருள் யார்யார்வாய்க் கேட்பினும் அப்பொருள்
மெய்ப்பொருள் காண்பது அறிவு

குறள் 423

Wisdom grasps the truth of whatever and by whomever said.

Kural 423

This Page Intentionally Left Blank

Preface

The initiation of biophysics in India is synonymous with the starting of the Department of Crystallography and Biophysics at the University of Madras, in 1953, under the leadership of Professor G.N. Ramachandran. Even today, nearly half-century later, there are only a handful of Universities in the country which have full-fledged biophysics departments. However, most life science curricula in the country now comprise a course in biophysics, owing to its increasing importance and relevance to the study of any biological system. This book has been written to fill a long-standing lacuna of textbooks in biophysics, suitable for students at the senior undergraduate and postgraduate level in various biological disciplines, such as biochemistry, molecular biology and even medicine.

Biophysics is an interdisciplinary subject, and its treatment varies with the background of both the author and the reader. For example, some aspects of theoretical biophysics are indistinguishable from pure mathematics. On the other hand, modern biophysics has made profound contributions to subjects normally considered a part of classical biology, such as evolution. In this book we have not assumed any knowledge of higher mathematics on the part of the reader as a prerequisite. We have emphasized an understanding of the underlying physical principles of the biological phenomena. Many of the explanations may therefore appear too elementary and too verbose to sophisticated readers. We believe however that the 'great divide' between students of the physical sciences who are mathematically fluent, and those of the biological sciences who are less so, is too deep to be bridged with one easy leap in a single textbook. Thus the present book does not set out to make biophysicists out of all biologists. However we believe it will cater to the needs of most biologists, who stop their acquaintance with mathematics at school. In addition, chemists and physicists may also find it useful as an introduction to the subject. The teaching notes of the authors form the skeleton of the book and this was fleshed out on the basis of several years of classroom experience.

Chapter 1 gives an introduction to laws of physics and chemistry and is designed to reinforce and enhance slightly the school-level understanding of the subject, which students with a background of purely biological science would have. This chapter covers quantum mechanics, molecular orbital theory, essentials of weak and strong molecular interactions, stereochemistry, fundamentals of thermodynamics and radiation biophysics.

Chapters 2 and 3 deal with separation techniques and physico-chemical techniques used to study

biomolecular structure. Strictly speaking this would be counted under physical biochemistry, rather than purely biophysics. The coverage is not exhaustive and is intended to give an overview of various techniques available to unravel the properties of biomolecules.

Chapters 4 to 9 describe some of the spectroscopy, microscopy, diffraction and computational techniques. These techniques are so powerful that every biophysicist must be armed with this knowledge so that it can be put to use when the need arises. These chapters aim at providing an understanding of the basic principles involved rather than thorough knowledge of the theory and practice of the techniques. The descriptions would enable most biologists to obtain an understanding of the relative merits and demerits of each technique, in order that the biochemical results, which one may come across in the general literature, become intelligible. X-ray crystallography and NMR spectroscopy have been dealt with in greater detail owing to their prominence in determining biomolecular structures.

Results obtained from the application of the techniques of structural biology are reviewed in Chapter 10. This chapter is illustrated by the largest number of diagrams. Nevertheless, in a book of this nature, in which one is always trying to keep costs low, it is not possible to include every picture the authors may feel necessary. Also we would have dearly liked to include colour illustrations.

The later chapters try and fill the gaps between biophysicists and biochemists. The treatment of Energy Pathways in Chapter 11 involves a lot of physical chemistry. Biomechanics in Chapter 12 and Neurophysics in Chapter 13 involve a lot of physiology. Though the advancements in the application of physical principles in these fields have been phenomenal they are probably too advanced and too detailed for the intended audience of this book.

We have of course referred to several books and journal articles, review papers, etc, while writing this book. At the end of the book we have included a chapter-wise list for further reading, which would amplify and extend the contents of each chapter. This list is of course neither exhaustive nor comprehensive. We have as far as possible selected books that would be available in most University libraries.

This book was written as a project sponsored and funded by the University Grants Commission, India. We take this opportunity to thank the UGC and all its officers for the invitation and support, financial and otherwise. We also thank the reviewers, whose corrections and suggestions we have tried to incorporate. Any errors that remain are, of course, ours. Our M.Sc., M.Phil and Ph.D. students have taught us (or have forced us to learn) much of the material presented in this book. We thank them all most sincerely. For actual help in preparing the book, we thank Mr A. Johnson and Mr. Prem Raj Bernard.

We would like to specially thank our respective family members for their enthusiastic support throughout. Writing this book has been a rewarding experience. But as we look back over it, we feel, with Thomas Hardy, ‘the more written, the more it seems remains to be written’.

VASANTHA PATTABHI
N. GAUTHAM

Contents

Preface *vii*

1. Laws of Physics and Chemistry **1**

1.1 Introduction *1*
1.2 Quantum Mechanics *1*
1.3 The Electronic Structure of Atoms *3*
1.4 Molecular Orbitals and Covalent Bonds *5*
1.5 Molecular Interactions *8*
1.5.1 Strong interactions *9*
1.5.2 Weak interactions *10*
1.6 Stereochemistry and Chirality *11*
1.6.1 Stereochemical nomenclature *13*
1.7 Thermodynamics *13*
1.7.1 Entropy *16*
1.7.2 Enthalpy *17*
1.7.3 The free energy of a system *17*
1.7.4 Chemical potential *18*
1.7.5 Oxidation-reduction potential *19*
1.8 Radioactivity *20*
1.8.1 Rate of radioactive decay *21*
1.8.2 Measurement of radioactivity *22*
1.8.3 Effects of radioactivity on matter *22*
1.8.4 Biological effects of radiation *23*
1.8.5 Applications of radio isotopes *23*

2. Separation Techniques **24**

2.1 Introduction *24*
2.2 Chromatography *24*
2.2.1 Column chromatography *26*

2.2.2 Thin layer chromatography *26*
2.2.3 Paper chromatography *26*
2.2.4 Adsorption chromatography *27*
2.2.5 Partition chromatography *27*
2.2.6 Gas liquid chromatography (GLC) *28*
2.2.7 Ion exchange chromatography *28*
2.2.8 Molecular exclusion chromatography *29*
2.2.9 Affinity chromatography *30*
2.3 Electrophoresis *32*
2.3.1 Moving boundary electrophoresis *32*
2.3.2 Zone electrophoresis *33*
2.3.3 Low voltage electrophoresis *33*
2.3.4 High voltage electrophoresis *33*
2.3.5 Gel electrophoresis *34*
2.3.6 Sodium dodecyl sulphate poly acrylamide gel electrophoresis (SDS-PAGE)
2.3.7 Iso electric focussing *35*
2.3.8 Continuous flow electrophoresis *35*

3. Physico-Chemical Techniques to Study Biomolecules **37**
3.1 Introduction *37*
3.2 Hydration of Macromolecules *38*
3.3 Role of Friction *38*
3.4 Diffusion *39*
3.5 Sedimentation *41*
3.6 The Ultracentrifuge *43*
3.7 Viscosity *46*
3.8 Rotational Diffusion *48*
3.8.1 Flow birefringence measurements *49*
3.8.2 Electric birefringence *50*
3.9 Light Scattering *51*
3.10 Small Angle X-ray Scattering *54*

4. Spectroscopy **58**
4.1 Introduction *58*
4.2 Ultraviolet/Visible Spectroscopy *59*
4.3 Circular Dichroism (CD) and Optical Rotatory Dispersion (ORD) *61*
4.4 Fluorescence Spectroscopy *65*
4.5 Infrared Spectroscopy *67*
4.6 Raman Spectroscopy *71*
4.7 Electron Spin Resonance *74*

5. Light Microscopy **77**
5.1 Introduction *77*
5.2 Elementary Geometrical Optics *77*
5.3 The Limits of Resolution *79*
5.4 Different Types of Microscopy *82*
5.4.1 Bright field microscopy *82*

5.4.2 Dark field microscopy *82*
5.4.3 Phase contrast microscopy *82*
5.4.4 Fluorescence microscopy *83*
5.4.5 Polarising microscopy *84*

6. Electron Microscopy **86**
6.1 Introduction *86*
6.2 Electron Optics *86*
6.3 The Transmission Electron Microscope (TEM) *87*
6.4 The Scanning Electron Microscope (SEM) *88*
6.5 Preparation of the Specimen for Electron Microscopy *90*
6.6 Image Reconstruction *91*
6.7 Electron Diffraction *92*
6.8 The Tunnelling Electron Microscope *92*
6.9 Atomic Force Microscope *93*

7. X-ray Crystallography **95**
7.1 Introduction *95*
7.2 Crystals and Symmetries *95*
7.3 Crystal Systems *96*
7.4 Point Groups and Space Groups *97*
7.5 Growth of Crystals of Biological Molecules *99*
7.6 X-ray Diffraction *103*
7.7 X-ray Data Collection *105*
7.8 Structure Solution *106*
7.8.1 The heavy atom or multiple isomorphous replacement (MIR) method *107*
7.8.2 Direct methods *108*
7.8.3 Molecular replacement *108*
7.8.4 The anomalous scattering technique *108*
7.9 Refinement of the Structure *108*
7.9.1 Least squares methods *108*
7.9.2 Constrained-restrained refinement *109*
7.9.3 Computer graphics in refinement *110*
7.10 Note on the Resolution of an X-ray Structure *111*

8. NMR Spectroscopy **112**
8.1 Introduction *112*
8.2 Basic Principles of NMR *112*
8.3 NMR Theory and Experiment *114*
8.4 Classical Description of NMR *115*
8.5 NMR Parameters *117*
8.5.1 Chemical shift *117*
8.5.2 Intensity *118*
8.5.3 Line width *119*
8.5.4 Relaxation parameters *119*
8.5.5 Spin-spin coupling *120*
8.6 The Nuclear Overhauser Effect *122*

8.7 NMR Applications in Chemistry *122*
8.7.1 Chemical shift *122*
8.7.2 Spin-spin coupling *123*
8.7.3 ^{13}C NMR *124*
8.8 NMR Applications in Biochemistry and Biophysics *124*
8.8.1 Concentration measurement *125*
8.8.2 pH titration *125*
8.8.3 Conformation of biomolecules *125*
8.8.4 Two-Dimensional NMR *126*
8.8.5 Determination of macromolecular structure *128*
8.9. NMR in Medicine *128*

9. Molecular Modelling **131**
9.1 Introduction *131*
9.2 Generating the Model *131*
9.2.1 Building the structure of H_2O_2 *132*
9.2.2 Building protein structure *133*
9.2.3 Building nucleic acid structure *136*
9.2.4 Displaying and altering the generated model *137*
9.3 Optimising the Model *138*

10. Macromolecular Structure **144**
10.1 Introduction *144*
10.2 Nucleic Acid Structure *144*
10.2.1 The chemical structure of nucleic acids *145*
10.2.2 Conformational possibilities of monomers and polymers *147*
10.2.3 The double helical structure of DNA *148*
10.2.4 Polymorphism of DNA *151*
10.2.5 DNA supercoiling and unusual DNA structures *154*
10.2.6 The structure of transfer RNA *158*
10.3 Protein Structure *161*
10.3.1 Amino acids and the primary structure of proteins *161*
10.3.2 The peptide bond and secondary structure of proteins *161*
10.3.3 Tertiary structure—supersecondary and domain structure *169*
10.3.4 Quaternary structure *170*
10.3.5 Virus structure *172*

11. Energy Pathways in Biology **174**
11.1 Introduction *174*
11.2 Free Energy *174*
11.3 Coupled Reactions *176*
11.4 Group Transfer Potential *177*
11.5 Role of Pyridine Nucleotides *178*
11.6 Photosynthesis *179*
11.6.1 Phostosystem I *183*
11.6.2 Photosystem II *185*
11.6.3 Photophosphorylation and carbon fixation *186*
11.7 Energy Conversion Pathways *187*

11.7.1 Oxidation *187*
11.7.2 Glycolysis *187*
11.7.3 The Krebs cycle *188*
11.7.4 The respiratory chain *190*
11.8 Membrane Transport *194*
11.8.1 Active transport *195*
11.8.2 Chemi-osmotic theory—passive transport *196*

12. Biomechanics 199
12.1 Introduction *199*
12.2 Striated Muscles *199*
12.2.1 Contractile proteins *201*
12.3 Mechanical Properties of Muscles *202*
12.3.1 Contraction mechanism *203*
12.3.2 Role of Ca^{2+} ions *204*
12.4 Biomechanics of the Cardiovascular System *205*
12.4.1 Blood pressure *205*
12.4.2 Electrical activity during the heartbeat *207*
12.4.3 Electrocardiography *208*

13. Neurobiophysics 210
13.1 Introduction *210*
13.2 The Nervous System *210*
13.2.1 Synapse *212*
13.3 Physics of Membrane Potentials *213*
13.3.1 Membrane potential due to diffusion *213*
13.3.2 Voltage Clamp *215*
13.4 Sensory Mechanisms—The Eye *217*
13.4.1 The visual receptor *218*
13.4.2 Electrical activity and visual generator potentials *223*
13.4.3 Optical defects of the eye *224*
13.4.4 Neural aspects of vision *224*
13.4.5 Visual communications, bioluminescence *226*
13.5 Physical Aspects of Hearing *227*
13.5.1 The Ear *227*
13.5.2 Elementary acoustics *228*
13.5.3 Theories of hearing *230*
13.6 Signal Transduction *230*
13.6.1 Mode of transport *231*
13.6.2 Signal transduction in the cell *231*

14. Origin and Evolution of Life 232
14.1 Introduction *232*
14.2 Prebiotic Earth *233*
14.3 Theories of Origin and Evolution of Life *234*
14.4 Laboratory Experiments on the Formation of Small Molecules *235*
14.4.1 Prebiotic synthesis of amino acids—Miller and Urey's experiment *235*
14.4.2 Prebiotic synthesis of purines, pyrimidines and nucleosides *237*

14.4.3 Sugars *237*
14.4.4 Nucleotide synthesis *238*
14.4.5 Optical activity *239*
14.4.6 Polymerisation *240*

FURTHER READING **243**

INDEX **247**

1

Laws of Physics and Chemistry

1.1 Introduction

Physics is an exact science. Biology, on the other hand, can be classified as a descriptive science. However, in recent times, the latter is also being transformed into a more exact science with the advent of molecular biology and molecular biophysics. One aspect of this transformation originates from the applications of physics to physiology. Physical laws or concepts such as mechanics, hydrodynamics, optics, electrodynamics and thermodynamics are used to explain physiological observations like muscle contraction, neural communication, vision, etc. A second and more fundamental aspect of the transformation emerged from the search for universal principles governing the world around us. Almost from the time man was able to think he has been wondering as to what makes him and the living world distinct from the rest of the environment. By the beginning of twentieth century it was realized that the laws of physics and chemistry, which were applied to non-living things, could equally well explain forces controlling biology, and that no new fundamentally different principles were necessary to explain the organisms and interactions which make up the living world. In the hundred years that have passed since then, this concept has become stronger and today nobody thinks it necessary to invoke any special physical or chemical forces or laws in the study of biology. This chapter will describe, in an elementary way, some of these basic concepts and laws of modern physics and chemistry, in particular those which are directly relevant to biology.

1.2 Quantum Mechanics

By the end of the nineteenth century, physical science had arrived at a description of the external world in terms of two distinct types of entities. On the one hand were particles and particulate matter to which the laws of mechanics, as enunciated by Newton and developed by several others, were wholly applicable. Concepts such as mass, momentum, kinetic energy, etc. were descriptive of particles. On the other hand were radiation and other wave like phenomena to which the laws of wave

propagation and other optical laws were applicable. Phenomena such as diffraction and interference could be explained in terms of the wave-like picture.

It was Max Planck who first suggested that there need not be a complete separation between the concept of a particle and the concept of a wave. In order to explain the observed results when the intensity of radiation from a perfectly radiating body (i.e. a so-called 'black body') was measured as a function of the wavelength, Planck proposed that electromagnetic radiation was emitted or absorbed not continuously but in small packets, or quanta. The energy contained within each quantum was given by the expression

$$E = h\nu$$

where ν is the frequency of the radiation and $h\ (= 6.625 \times 10^{-34}$ joule seconds) is a universal constant called Planck's constant. The black body could only absorb or emit integral multiples of this quantity, never any fraction of a quantum. Einstein extended this idea to say that not only is radiation absorbed or emitted in packets but it also propagates only as packets, called photons, each containing an amount of energy given by the Planck expression. Thus, radiation, which was thought to be wavelike, now could be treated as a particle and the same entity could both diffract as well as possess momentum. The French scientist De Broglie suggested that the same type of wave-particle duality could be applied to entities hitherto thought to be particles. Thus, he said, electrons and other sub-atomic particles could be thought of also as wave packets, with a wavelength given by

$$\lambda = h/p$$

where p is the momentum of the particle. In fact this equation is applicable not only to sub-atomic particles but to all matter. However Planck's constant is so small that the wave nature of matter becomes observable only when the mass of the particle, and hence its momentum, is also very small. When this happens, then according to De Broglie's theory, a particle such as an electron would not only possess mass and momentum but would also show diffraction and other wave-like effects. Hundreds of thousands of experiments performed all over the world since the proposals were made have firmly established the truth of the wave-particle duality of the physical world. Quantum mechanics is now accepted as the only correct description of physics. However for practical calculations on macroscopic objects like cricket balls and space rockets it is sufficient to use Newton's laws. For calculations of atomic level phenomena we use quantum mechanics, chiefly as developed by Schrodinger and Heisenberg.

Schrodinger introduced a wave equation that replaces the classical equations of motion. The solution to this wave equation is a 'wave function' which is a function of space and time. The physical state of any system such as an electron (or an atom or a molecule or a system of molecules) can be described completely by the corresponding wave function. All physically meaningful and measurable quantities such as the energy, the momentum, the position and so on can be extracted from the wave function. The wave function is usually represented as Ψ and is obtained as a solution of the following 'eigenvalue' equation

$$H\Psi = E\Psi$$

where E is the characteristic energy of the system or the energy 'eigenvalue' and H the operator called Hamiltonian operator, which operates upon the characteristic wave function or the 'eigenfunction' Ψ to yield the energy values. A physical way of understanding Ψ is as a probability amplitude. In other words, since Ψ is a function of the coordinates x, y and z and time t, the $|\Psi|^2$ gives the probability of finding the system at the point x, y, z at time t. Written explicitly, the Schrodinger wave equation has the following form

$$\frac{d^2\psi}{dx^2} - \frac{8\pi^2 m}{h}(E - V)\,\psi = 0$$

where m is the mass of the particle and V the expression for the potential energy. This is a one-dimensional equation, independent of time. To solve this equation one has to, first of all, define the appropriate expression for the potential energy V, which will depend on the problem studied. When this expression is inserted into the Schrodinger wave equation, the differential equation so obtained can be solved to find Ψ and E. In three dimensions the Schrodinger equation becomes

$$\frac{d^2\psi}{dx^2} + \frac{d^2\psi}{dy^2} + \frac{d^2\psi}{dz^2} - \frac{8\pi^2 m}{h}(E - V)\,\psi = 0$$

The time dependent form of the equation is as follows:

$$-\frac{h^2}{8\pi^2 m}\left(\frac{d^2\psi}{dx^2} + \frac{d^2\psi}{dy^2} + \frac{d^2\psi}{dz^2}\right) = \frac{ih}{2\pi}\frac{d\psi}{dt}$$

where i is the imaginary root.

This scheme of calculations involving the wave equation is called *wave mechanics*. It is complemented by another, equivalent scheme developed by Heisenberg and Born, called *matrix mechanics*. The starting point for these calculations was the discovery made by Heisenberg that there exists a physical limit to the accuracy with which one can simultaneously determine both the position and the velocity of a particle. This is due to the fact that any measurement necessarily implies that the particle under observation is being subjected to a disturbance, at the very least by a photon of light. If the particle is small, then the act of observation itself leads to uncertainty in the position. (If the particle were large, then since it is extended it cannot be said to possess a very precise location). Heisenberg stated these facts in the form of the Uncertainty Principle, which declares that no two canonically conjugate observables can be measured simultaneously with arbitrary precision. Position and momentum are canonically conjugate variables. So are *energy* and *time*. Another such pair is *angular momentum* and *angular position*.

Both matrix mechanics and wave mechanics may be used to arrive at the physically meaningful and measurable variables that characterize any physical system. Which scheme of calculation is to be used depends on the actual information that is sought. The most spectacular and immediate success of quantum mechanics was in the theoretical description of the structure of atoms.

1.3 The Electronic Structure of Atoms

Rutherford's model of the atom described it as consisting of a central nucleus with a positive electric charge Ze, and Z electrons each with a negative electric charge e revolving around it. Z is the atomic number of the atom. In order to resolve certain inconsistencies in this picture, Bohr proposed an altered model, which was later shown by quantum mechanical calculations to be the correct one. Bohr postulated that the electrons, which revolve round the central nucleus, occupy only certain fixed orbits. These orbits are such that the associated angular momentum is always an integral multiple of $h/2\pi$. They are called the stationary quantum states of the atom and can be calculated as the solutions of the Schrodinger wave equation (Figure 1.1). When the electrons are in one of the stationary states, no radiation occurs, in spite of the fact that according to classical electrodynamics one would expect such radiation due to the accelerated motion of the charged electrons. Bohr also postulated that radiation is emitted or absorbed only when electrons jump from one stationary state (also called an orbital) to another. The difference in energy between two states is given by

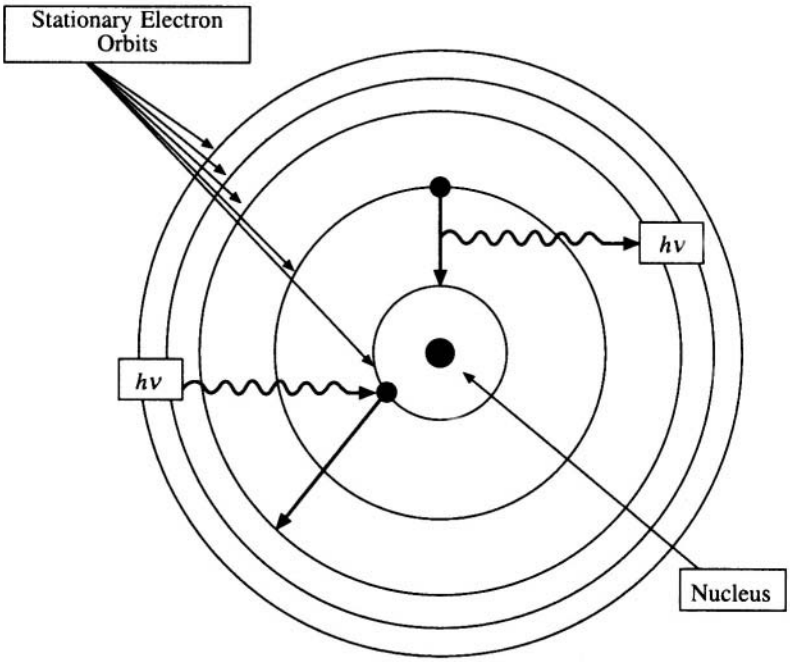


Fig. 1.1 Stationary states in the Bohr atom

$$\Delta E = h\nu = hc/\lambda$$

where λ is the wavelength of the radiation, and c the velocity of light. This equation determines the wavelength of the radiation emitted or absorbed (Figure 1.2). The motion of each of the electrons in the atom is described by a wave function ψ, known also as an atomic orbital. A combination of the wave functions of all the electrons leads to the wave function Ψ for the atom as a whole. Four quantum numbers are used to describe the orbital of an electron. These are (a) principal quantum number n, (b) angular momentum quantum number l, (c) magnetic quantum number m_l, and (d) spin quantum number m_s. The principal quantum number specifies the total energy of the electron. The angular momentum associated with the orbital motion of the electrons is described by the angular

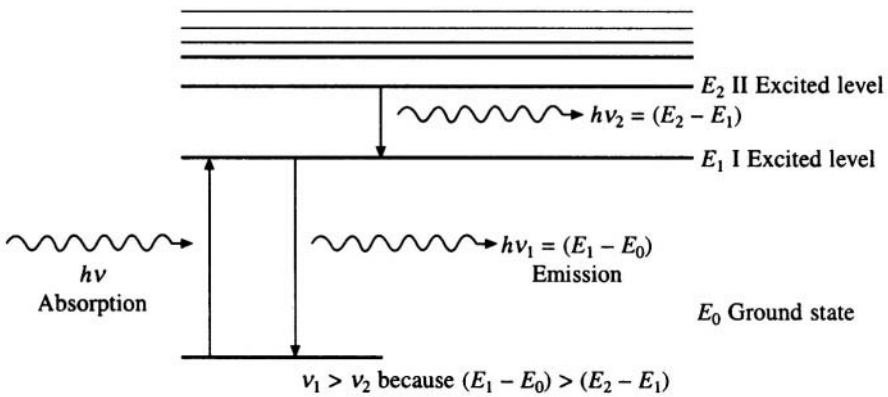


Fig. 1.2 Energy levels and the transitions between them

momentum quantum number. The principal quantum number can assume only integral values (n = 1, 2, 3,…). There can be more than one electron with the same principal quantum number, though for each electron at least one of the quantum numbers must be different. Electrons with the same value of the principal quantum number n constitute an electron shell. These shells are designated *k,l,m,...* etc., corresponding to n = 1,2, 3,… etc. The angular momentum quantum number l can only take on the values 0, 1, 2, 3, ... $n - 1$. In the presence of a magnetic field the orientation of the vector representing the angular momentum is described by the magnetic quantum numbers m_l, which can have values 0, ±1, ±2,… ±l. Lastly, the spin angular momentum is due to the spin of the electron about its own axis. The spin vector can assume one of two possible orientations in the presence of a magnetic field and the corresponding values of m_s are +1/2 and –1/2. Pauli's Exclusion Principle states that no two electrons can have all the four quantum numbers the same. Two electrons may occupy the same space orbital. Then their spins must be anti-parallel to each other, i.e. if one of them has $m_s = +1/2$, the other has the value –1/2. The orbital with the lowest energy is known as the ground state. For an atom with many electrons the ground state is when the electrons are arranged in the lowest energy orbitals consistent with the exclusion principle. Other states are called excited states.

When an electron (or the atom as a whole) makes a transition from one quantum state to another, energy is either absorbed or emitted. The radiated energy is equal to the difference in energy levels of the initial and final states. The frequency v of the absorbed or emitted radiation can be written as

$$v = \Delta E/h$$

where ΔE is the difference in the energies of the two levels. On atoms with more than one electron, the spin and orbital angular momenta of the individual electrons in an atom combine to produce a net angular momentum for the atom as a whole. Thus an atom with two electrons can have a total spin equal to zero when the electrons are paired (i.e. they have anti-parallel spins) or equal to one when they are unpaired (parallel spins). However, in the latter case the electrons must occupy different orbitals according to Pauli's Exclusion Principle. The orbitals are designated *s, p, d,f etc.* according to the value of the angular momentum quantum number. The 1*s* orbital is the lowest energy orbital with n = 1, l = 0 and $m_l = 0$. This orbital can be occupied by two electrons with spins +1/2 and –1/2 respectively. The orbital with the next higher energy value is the 2*s* orbital, followed by the 2*p* orbital, and so on (Table 1.1).

Table 1.1 Electron distribution in atomic orbitals of the first three shells

Atomic Shells	n	l	m_l	m_s	Orbital Designation	Number of Electrons Accommodated
K	1	0	0	+1/2, –1/2	1*s*	2
L	2	0	0	+1/2, –1/2	2*s*	2
		1	+1, 0, –1	+1/2, -1/2	2*p*	6
M	3	0	0	+1/2, –1/2	3*s*	2
		1	+1, 0, –1	+1/2, –1/2	3*p*	6
		2	+2, +1, 0			
			–1, –2	+1/2, –1/2	3*d*	10

1.4 Molecular Orbitals and Covalent Bonds

An exact analytical solution of the Schrodinger equation is possible only for simple systems. In order to obtain the energy eigenvalues and eigenfunctions of systems which have more than one electron,

a number of approximate methods are available. The molecular orbital method is one such approximate method wherein wave functions for the electrons in a molecule are constructed from atomic orbitals, i.e., the wave functions for single electrons in the atoms that form the molecule. When a linear combination of atomic orbitals is used to describe a molecular orbital then the method is called MO-LCAO approximation (Linear Combination of Atomic Orbitals).

Consider, for example, a hydrogen molecule where two hydrogen atoms A and B come together to form a molecule. When the atoms are far apart their electrons occupy the 1*s* orbital ($n= 1, l = 0$). But when they are brought together to form a molecule, the molecular orbital is given by

$$\Psi_{MO} = C_1\phi_A(1s) + C_2\phi_B(1s)$$

where ϕ_A and ϕ_B are atomic orbitals and $C_1 = \pm C_2$. Hence, we get

$$\Psi_1 = C_1\phi_A(1_s) + C_1\phi_B(1s)$$

$$\Psi_2 = C_1\phi_A(1s) - C_1\phi_B(1s)$$

Ψ_1 is called an even orbital and Ψ_2 the odd orbital. In case of an even orbital the electrons are shared by the two nuclei while in an odd orbital there is no electron density between the nuclei. This orbital has higher energy (Figure 1.3). The even orbitals are bonding orbitals and the odd orbitals are anti-bonding orbitals. Note that if the hydrogen molecule is in the anti-bonding state, this does not mean that the electrons or the nuclei will fly apart. On the contrary, the bond between the two atoms remains, except that this is now unstable and of higher energy, and the molecule tends to make a transition to the lower energy bonding state. When the hydrogen molecule is formed, the bonding orbital is occupied by two electrons with anti-parallel spins. When the molecule is excited one of the electrons jumps to the anti-bonding orbital. As in the case of atoms, molecular orbitals also have associated quantum numbers. In a diatomic molecule like hydrogen, n is the principal quantum number and l is the quantum number that gives the angular momentum component along the inter-nuclear axis. $l = 0, 1, 2\ldots$ are called $\sigma, \pi, \delta\ldots$ orbitals, analogous to the quantum number l in atoms.

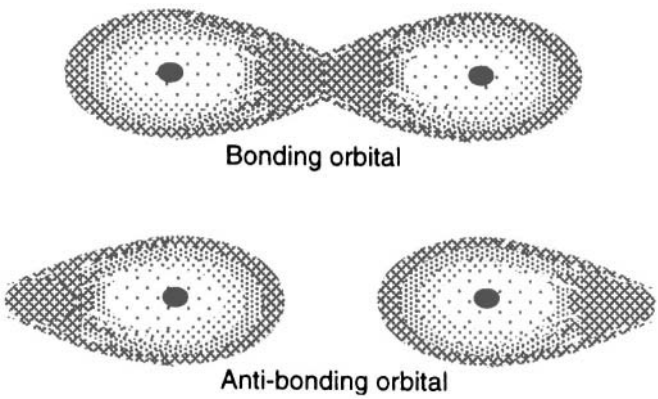


Fig. 1.3 Bonding and anti-bonding orbitals in a hydrogen molecule

In diatomic molecules, σ orbitals are symmetric with respect to the inter-nuclear axis and π orbitals are anti-symmetric. The higher orbitals are more complex (Figure 1.4). In multi-atomic molecules the atomic orbitals first combine to form a hybrid orbital before a molecular orbital is formed. Carbon, which is a pivotal atom in biology, is a classic example for the formation of hybrid orbitals. The electronic configuration of carbon is $1s^2 2s^2 2p^2$ (i.e. two electrons in the 1*s* orbital, 2 electrons in the

2*s* orbital and 2 electrons in the 2*p* orbital accounting for the six electrons in carbon). However a lower energy state is achieved if one of the electrons from the 2*s* level is promoted to a 2*p* state. Thus, the atom has four unpaired electrons—one in 2*s* orbital and three in 2*p* orbital. These four orbitals combine to form hybrid orbitals. The stable way of doing this is to combine the 2*s* orbital with each component p_x, p_y and p_z of the *p* orbital in the following way

$$\Psi_1 = 1/2\ (\phi_s + \phi_{px} + \phi_{py} + \phi_{pz})$$

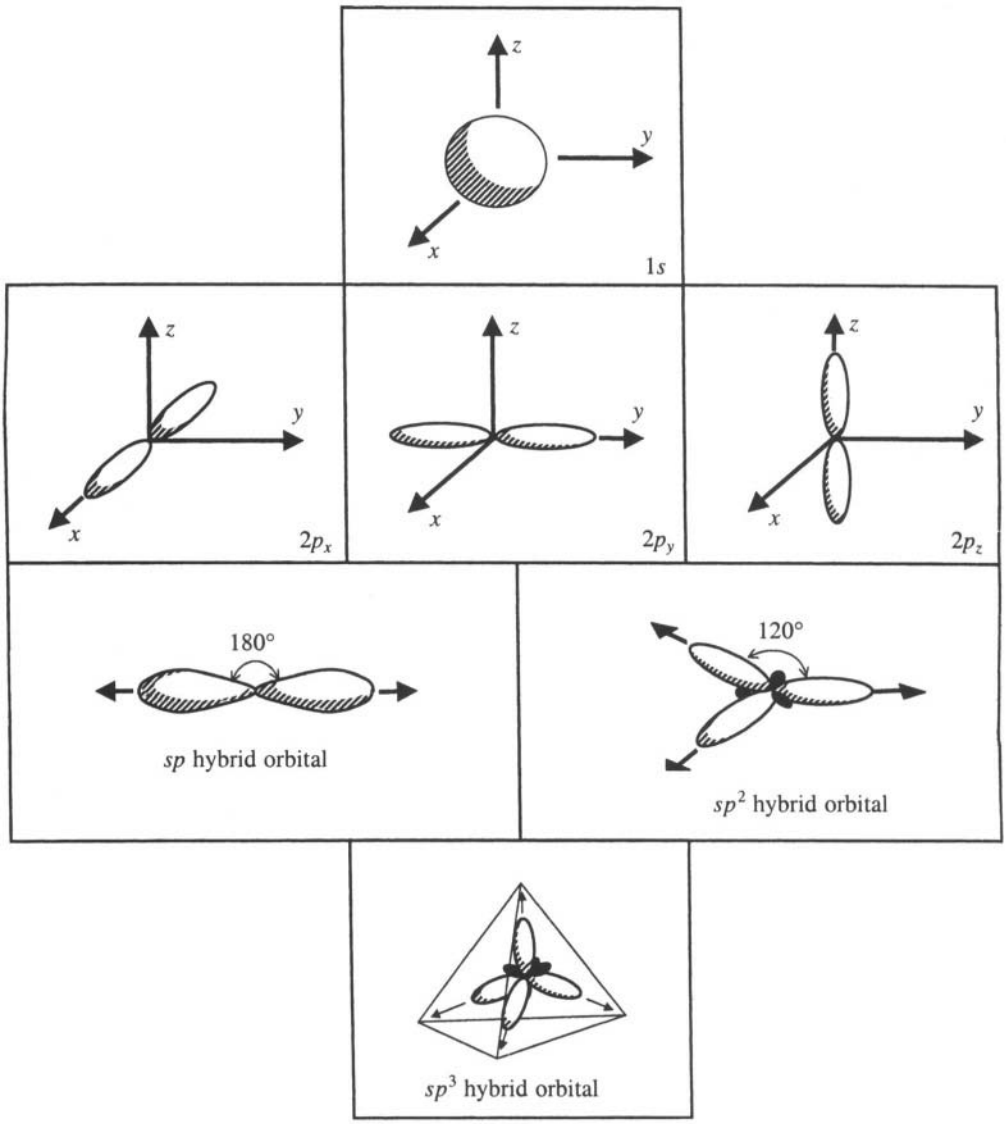


Fig. 1.4 Pure and hybrid atomic orbitals

$$\Psi_2 = 1/2\ (\phi_s + \phi_{px} - \phi_{py} - \phi_{pz})$$

$$\Psi_3 = 1/2\ (\phi_s - \phi_{px} + \phi_{py} - \phi_{pz})$$

$$\Psi_4 = 1/2\ (\phi_s - \phi_{px} - \phi_{py} + \phi_{pz})$$

(*p* orbitals have a butterfly-like appearance with their nodal point at the nucleus. The p_x, p_y and p_z orbitals are at right angles to each other, see Figure 1.4). The resultant hybrid is known as the sp^3 orbital and points towards the four corners of a tetrahedron (Figure 1.4). When these hybrid orbitals combine with the 1*s* orbitals of four hydrogens, a methane molecule is formed through four σ bonds (covalent bonds). The strongest bonds are formed when the overlap of the orbitals is a maximum.

Other schemes of hybridization are also possible. For example, sp^2 hybrids, also known as trigonal hybrids, are formed when 2*s*, $2p_x$, and $2p_y$ orbitals are mixed. This leads to three coplanar orbitals making an angle of 120° with each other (Figure 1.4). The p_z orbital is perpendicular to the plane. When a molecule such as ethylene (Figure 1.5) is formed, the sp^2 orbitals fuse with each other and with the 1s orbital of hydrogen to form localized σ orbitals. The p_z orbitals, which extend below and above the plane, also fuse together to form π orbitals. In benzene the sp^2 atomic orbitals of the six carbon atoms, fuse together to form six σ orbitals in a closed ring. The six electrons in these orbitals are not localized to a particular atom but move freely in the ring. The orbitals are therefore known as delocalized orbitals and the molecule is said to have conjugated double bond or resonance structure. If two 2*p* orbitals combine to form a π molecular orbital it could be a bonding orbital (π) or an anti-bonding orbital (π^*). Transitions from one molecular orbital to another, for example from π to π^* or from *n* to π^* are called $\pi\pi^*$ or $n\pi^*$ transitions respectively. As in the case of atoms, electronic transitions in molecules also lead to emission or absorption of light. Owing to the differences in the energy levels, $\pi\pi^*$ transitions are about 100 times more intense than $n\pi^*$ transitions.

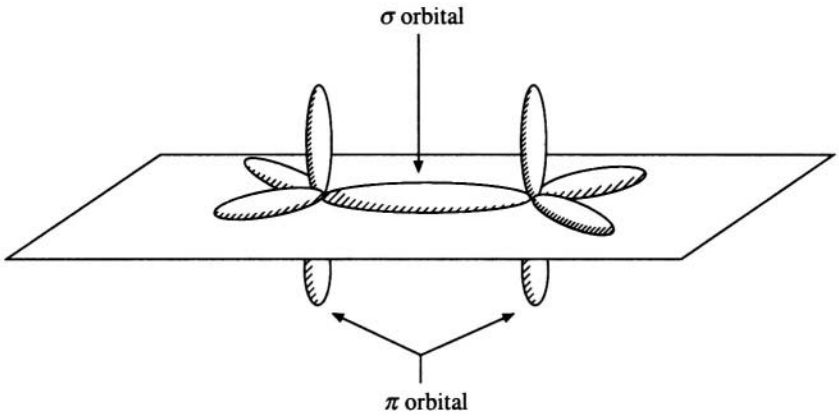


Fig. 1.5 The σ and π molecular orbitals

1.5 Molecular Interactions

The structure of a biological molecule determines its function. In turn, the forces between the atoms in a molecule determine its structure. The interactions between the atoms in the molecule are classified as strong or weak, depending on whether or not the interaction can be disrupted by weak forces like thermal motion. (It is customary to measure the strength of an interaction by the energy required to disrupt it.) Thus, for example, the primary structure of biological macromolecules is made of strong interactions such as the covalent peptide bond. Higher order structures like secondary, tertiary and

quaternary structures are governed by weak forces and can, therefore, be disturbed by a relatively small increase in temperature, or a change of pH, etc. Strong interactions are implicated mainly in the formation of the chemical structure, and, to some extent, also in the formation of the molecular structure. Weak interactions, on the other hand, not only help to determine the three-dimensional structure, but also are involved in the interactions between different molecules. Any interaction within a molecule or between molecules can be understood as a sum of the interactions between pairs of atoms. When two atoms come close to each other a potential energy is associated with the system. This potential energy is a minimum (largest negative value) when the force of attraction between the atoms equals the force of repulsion, and is zero when the two atoms are separated by an infinitely large distance (Figure 1.6). The system as whole tries to attain an equilibrium minimum energy state and the final equilibrium distance between the two interacting atoms is a resultant of all the different forces, which act on each of them. Some of these forces are described below.

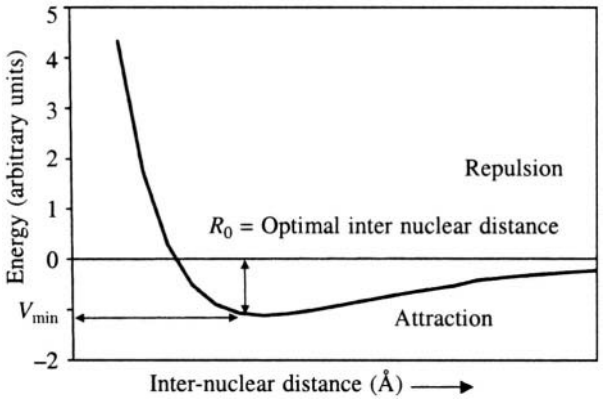


Fig. 1.6 Potential energy between two atoms plotted as a function of the inter-nuclear distance

1.5.1 Strong interactions

Covalent bonds, ionic bonds and metallic bonds are classified in this category. An ionic bond is formed due to electrostatic attraction between two oppositely charged ions. The structure of sodium chloride is an example of this type of bonding. In a crystal of common salt, every sodium ion is surrounded by six chlorine ions, and vice versa (Figure 1.7). Note that, strictly speaking, the entire crystal is one large 'molecule' of sodium chloride, consisting of sodium and chlorine atoms held together by ionic forces, and an isolated NaCl molecule has no independent existence[1]. In such ionic compounds an electron is transferred from an electropositive to an electronegative element in order to achieve a stable electron configuration. In NaCl, sodium looses one electron and becomes an ion, Na^+, with a single positive charge but a stable electronic configuration $1s^2 2s^2 2p^6$, with completely filled orbitals. In a similar fashion, chlorine acquires one electron to attain the same stable neon-like

[1]When Lawrence Bragg determined the structure of NaCl by X-ray crystallography and suggested that NaCl did not exist as a single molecule by itself, he was subjected to much ridicule by chemists for daring to suggest that they give up their cherished prejudices. They said that as a physicist he should stick to subjects he understood, rather than speak what they thought was obvious nonsense. History, of course, proved Bragg right.

configuration and becomes Cl^-, again an ion, with a single negative charge. The positive ion and the negative ion come together to form a strong ionic bond and thence the crystal of sodium chloride.

Of the other strong interactions, covalent bonds have already been described. Metallic bonds are not relevant in biological systems and will not be further dealt with here.

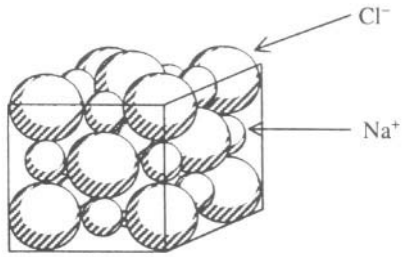


Fig. 1.7 Structure of crystalline sodium chloride salt with Na^+ and Cl^- ions placed alternately in a face-centered cubic cell.

1.5.2 Weak interactions

Hydrogen bonds and van der Waals interactions are classified in this category. Van der Waals forces (also known as residual forces) are weak forces, which hold together inert elements and molecules. Most organic crystals, for example, are built of neutral molecules and this type of force plays an important role in their stability. Van der Waals forces act between all atoms and ions in all solids but the effect cannot be felt in the presence of strong interactions like covalent, ionic or metallic bonds. Van der Waals forces are basically electrostatic in nature in that they involve interactions between electric dipoles. When two bonded atoms have a difference in their ability to attract electrons (i.e. in their electronegativity), they form an electric dipole and hence possess a dipole moment. For example, in a hydroxyl group (OH) the oxygen atom draws the electron away from the hydrogen nucleus. There is thus a net (fractional) positive charge centered on the hydrogen nucleus, and a corresponding net (fractional) negative charge around the oxygen nucleus. In other words a dipole is formed (Figure 1.8). Note that the molecule as a whole is electrically neutral. The dipole moment may be a permanent feature arising from the molecular structure, if relatively electropositive atoms are bonded to relatively electronegative ones, and the dipoles of all the bonds do not cancel. In such a case a residual dipole moment may exist independent of the environment of the molecule. Temporary dipoles may also arise by induction due to, for example, the presence of an external electric field. Both permanent and temporary dipoles arise in solids in which the molecules are polar. But in the case of inert elements such as helium, neon, etc., and non-polar, symmetric molecules, such as methane, hydrogen, etc., London proposed that a dispersion effect would lead to momentary polarisation of the charge, and hence a dipole. Even in the case of a highly symmetrical system with no permanent dipole moment, the electronic distribution at any instant of time may not be perfectly symmetrical, due to the constant motion of the electrons. This lack of symmetry gives rise to an instantaneous dipole moment. This instantaneous dipole moment of an atom or a molecule will polarise the neighbouring atoms or molecules, and a momentary attractive force arises between the two. This adds up to give a net attractive van der Waals force due to London dispersion. There are thus three components of the van der Waals force, viz., the permanent dipoles in the molecule, the dipoles induced by an external electric field, and the ones induced by the London dispersion effect. The

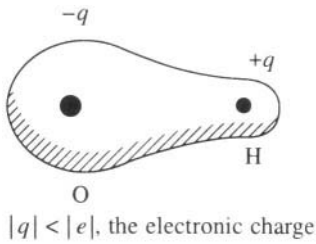


Fig. 1.8 The dipole moment in a hydroxyl group arising out of the difference in electronic density surrounding the two nuclei. The charge q is less than one full electronic charge.

relative contribution of each to the total force varies, and depends on the type of molecule. At the level of the individual atoms, others define the term 'van der Waals radius' as the distance of closest approach to the centre of the atom, in the presence of van der Waals forces alone. This radius can be experimentally determined by X-ray crystallography. In general, the van der Waals radius of an atom is larger than its covalent or ionic radius, showing that the effective radius of the atom within a molecule is different from (and less than) the distance of closest approach by other molecules or atoms.

Hydrogen bonds are weaker than covalent bonds but are stronger than van der Waals bonds. Hence the atoms connected by a hydrogen bond are separated by a distance (2.5 to 3.0 Å) greater than that of a covalent bond but shorter than if van der Waals forces alone were operative. Hydrogen bonds are generally found between two electronegative atoms. If *X* is an electronegative atom bonded to a hydrogen atom, it draws the electron of the hydrogen atom towards it, leaving the latter with a residual positive charge. The magnitude of this residual charge is less than one, since the electron is not completely stripped away from the nucleus, but only displaced more towards the nucleus of the atom *X*. The *H* ion, with a fractional positive charge, is now attracted by the electrons surrounding the second electronegative atom *Y*. As the hydrogen ion is almost devoid of its only electron, it can penetrate the electron cloud of atom *Y* till it is stopped by the repulsive forces of the nucleus of atom *Y*. Thus a bond is formed between atom *X* and atom *Y*, called the hydrogen bond. A hydrogen bond is generally written as *X–H ...Y* since the *H* ion is asymmetrically disposed with respect to *X* and *Y*. It is particularly associated with *X* and loosely bonded to *Y*. Hydrogen bonds are among the most important of the interactions between biological molecules and confer specificity to the recognition of one molecule by the other. A prime example of this is formation of the highly specific A.T and G.C base pairs in DNA, which depends almost solely on the number and nature of the hydrogen bonds between the pairs.

1.6 Stereochemistry and Chirality

The three-dimensional spatial arrangement of the atoms of a molecule is termed its stereochemistry. A molecule may be identified first by its chemical formula, then by its chemical structure, and finally by its molecular structure. For example, C_2H_6O is the chemical formula of ethyl alcohol. $CH_3–CH_2–OH$ is the chemical structure of ethyl alcohol. The molecular structure is determined by the three dimensional arrangement of the atoms in space. The word 'configuration' is used to describe the distinctive arrangement of various atoms or groups attached to a chiral centre (as described below). The word 'conformation' describes the different arrangements of atoms that are obtained when parts of the molecule are rotated about one of the bonds. A change in the configuration necessarily involves breaking at least one covalent bond and reforming it after a rearrangement of the atoms. A change in the conformation necessarily does not involve any breaking of covalent bonds. Groups that are connected by a single bond can undergo rotation leading to different relative orientations with respect to each other. For example, in ethyl alcohol, depending on the angle of rotation about the C–C bond, the relative orientations of hydrogen atoms and the hydoxyl group will vary. Energetically, the most favoured conformation about any bond is the 'staggered' one in which atoms attached at either end of the bond are at the largest possible distance from each other. An 'eclipsed' conformation increases steric contacts and is much less favourable. Arrangements of atoms or groups arrived at by rotation around a double bond are configurational isomers not conformational isomers. This is because the double bond does not allow free rotation. To produce the isomers, it becomes therefore necessary to break the double bond and then reform it after the rearrangement of the atoms. *cis-trans* isomers fall under this category and are observed in the case of the amino acid proline and in polypeptides. Linear

polypeptides consist almost exclusively of trans peptide units while cyclic peptides can have both *cis* and *trans* peptide units.

In some molecules the differences in the arrangement of the atoms may not affect many of the physical properties of the molecule such as melting point, density, refractive index, etc., but may affect the interaction of the different molecules with polarised light. Such molecules are said to be optically active and are called optical isomers. Optically active molecules rotate in different directions the plane of polarisation of the plane polarised light incident on them. Levo-rotatory molecules rotate the plane in the clockwise direction and dextro-rotatory molecules rotate it in the anticlockwise direction. These optically active molecules may differ in their biological behaviour. For example, only one of the optical isomers of aspargine is sweet to taste. Louis Pasteur showed that the difference in action between two optical isomers is due to the difference in molecular architecture, i.e. due to differences in the arrangement of the atoms in space. The two types of molecules are related by a mirror reflection and they cannot be superimposed. Objects of everyday life which are enantiomers (i.e. related by mirror symmetry) are the right hand and the left hand, righthanded and lefthanded screws, etc. The two alanine molecules shown in Figure 1.9 have the same number of atoms but differ in their stereochemistry and hence are known as stereoisomers. Compounds with four different substitutions at a carbon atom will exhibit optical activity. Such a carbon atom is generally known as an asymmetric carbon and a molecule containing such an atom is a chiral molecule. A mixture of two isomers, which have optical activity opposite to each other, is on the whole optically inactive since the effect of one is annulled by the other. Such a mixture is known as a racemic mixture.

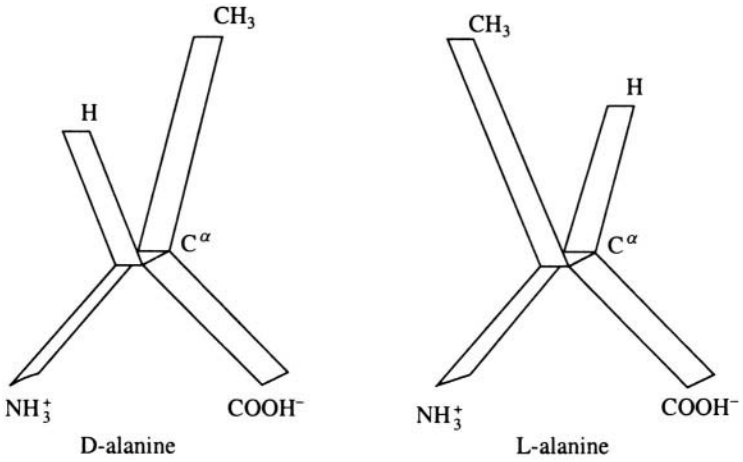


Fig. 1.9 A schematic representation of *D*-alanine and *L*-alanine

A pair of enantiomeric molecules will behave in the same way with respect to symmetrical environments or achiral chemical reagents. However, their interaction with other chiral molecules will differ. An example of this is the very specific interaction between a protein and its substrate. Stereoisomers not related by mirror symmetry are called diastereoisomers. Diastereoisomerism will occur in molecules with more than one chiral centre. Both configurational and conformational differences can lead to optical isomers— i.e. either enantiomers or diastereoisomers. Other than carbon, atoms like silicon with a valency of four can also act as chiral centres. Some molecules like hexahelicene, and polyglycine show optical activity though they have no chiral centre. This is due to the helical structures assumed by these molecules, which can be either right handed or left handed (Fig. 1.10).

All the major biological molecules, like amino acids, proteins, DNA, sugars and lipids, are optically active. Out of the two or more optical isomers available for these molecules only one kind is commonly found in biological systems. Optical activity and Life seem to be inseparable.

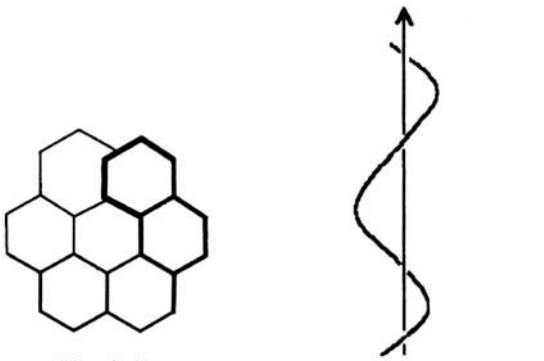


Fig. 1.10 Optically active molecules, but without a chiral centre

1.6.7 Stereochemical nomenclature

Molecules are three-dimensional entities. To represent them in two dimensions, i.e. on a piece of paper some standard methods are available. The most widely used are the Newman projection and the Fischer projection (Figure 1.11). In addition to these, there is another convention used to describe the arrangements of various groups around a chiral centre. This notation is known as the Cahn-Ingold-Prelog system and the enantiomers are called the [R] and the [S] configurations. The R–S convention is as follows: If the groups attached to the chiral centre are arranged according to the sequence rules (given below) as $a \rightarrow b \rightarrow c \rightarrow d$, then if the chiral centre is viewed from the side opposite to the lowest group (i.e. *d* in this case), and if $a \rightarrow b \rightarrow c$ occur in a clockwise direction, the molecule is in the [R] configuration. On the other hand if they occur in the anti-clockwise direction then it is in the [S] configuration (Figure 1.12). The sequence rules are as follows: (1) Atoms are arranged according to atomic number, e.g. Cl > O > N > C > H. (2) If the attached atoms are the same, then the next atom is used for ordering, e.g. $CH_3—CH_2—CH_2— > CH_3—CH_2— > CH_3—$. (3) If the substituents have atoms of the same rank then the one with more atoms of the highest rank gets preference, e.g. $(CH_3)_3—C— > (CH_3)_2—CH— > CH_3—CH_2—$ etc., and $CH_3—OH— > CH_3—CH_2—$ and so on. Since at each chiral atom there are two ways in which the attached groups can be arranged, if there are *n* chiral centers then there will be 2^n stereoisomers.

In the case of amino acids a different notation is used. In this molecule the C^{α} atom is the chiral centre. If we look down the $H—C^{\alpha}$ bond and if $CO \rightarrow R \rightarrow N$ is read in the clockwise direction, then it is the *L* configuration. If it is read in the anti-clockwise direction then it is the *D* configuration (Figure 1.9). All naturally occurring amino acids are in the *L* configuration.

1.7 Thermodynamics

Thermodynamics deals with energy, heat and work. Any system, physical or chemical or even biological, can be considered as a thermodynamical system and the laws of thermodynamics can be applied to it, if it utilizes energy in any form. To introduce some of the thermodynamic concepts, consider the

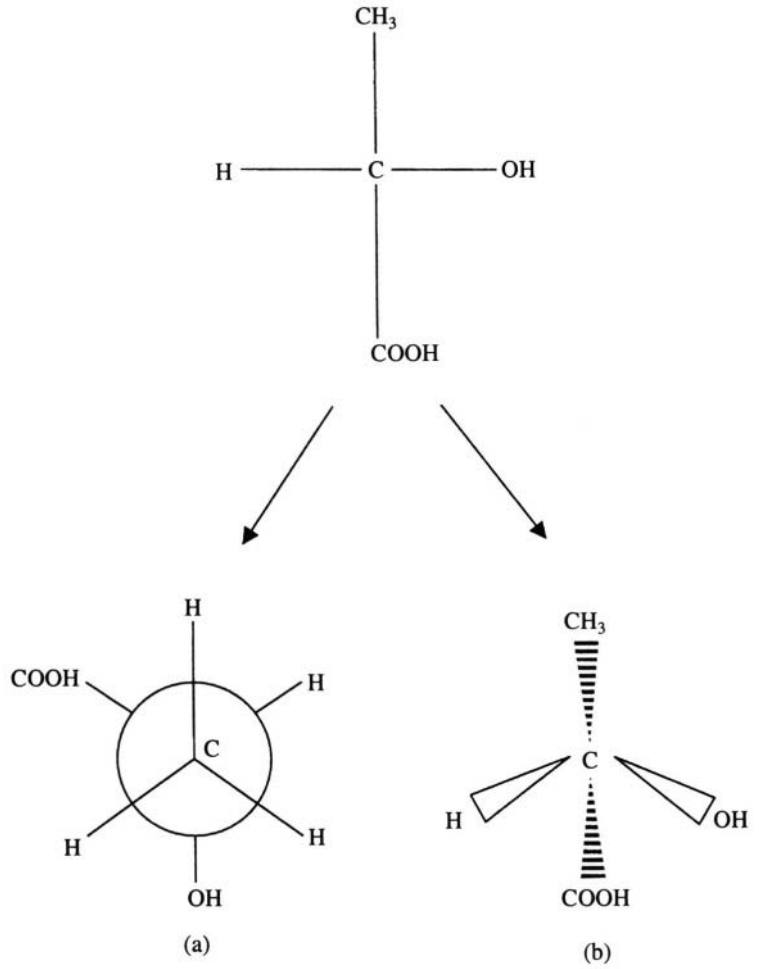


Fig. 1.11 (a) The Newman projection. (b) The Fischer convention

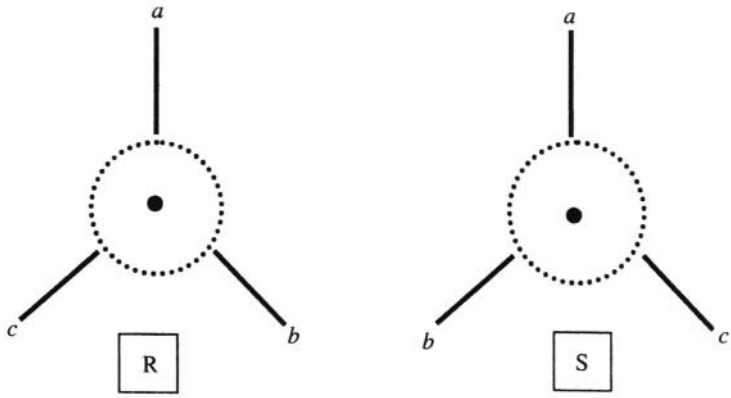


Fig. 1.12 The Cahn-Ingold-Prelog nomenclature system to distinguish R and S enantiomers

simple example of a glass of hot water on a table. If the glass is not disturbed in any way, the temperature of the water slowly decreases till it reaches the temperature of the room. This is achieved by transferring heat energy from the hot water to the environment. The glass of water is considered the system and the room is the environment. If T_s and T_e represent the temperature of the system and the environment respectively, then, if $T_s > T_e$, heat will flow from the system to the environment. If $T_s < T_e$, the flow of heat will be in the other direction. Heat energy Q is positive when heat flows from the environment into the system and is negative when the opposite happens. A system is said to be isolated when there is no exchange of energy or matter with the surroundings. When there is an exchange of energy but not of matter then the system is said to be closed. When both matter and energy can be freely exchanged with the environment, then the system is an open one.

In the previous example of the glass of water, we consider that heat, i.e. energy, but not matter, can be exchanged between the glass and the room. The system is a closed one and since initially the water is hotter than its surroundings, heat flows from the system to the environment, until $T_s = T_e$. At this point the system is said to be in equilibrium with the surroundings. The equilibrium in this case is a stable one and the temperatures remain equal, unless an external source (or sink) of heat disturbs one of them. The process by which the glass of hot water has attained this equilibrium with its environment is called an irreversible process. A process can be reversible or irreversible depending on whether or not it can trace back the same path to reach the initial state. A reversible process (considering the system and environment together) goes through infinitely small changes at an infinitely slow pace, and is always at equilibrium. Natural or biological processes almost always take place spontaneously and transfer energy from one part of the system to another in a short time, and hence are not thermodynamically reversible processes.

Thermodynamics requires a measurement of energy. The units of energy are calories or joules, depending on whether it is expressed as heat or as work. One calorie is defined as the amount of heat required to raise the temperature of 1 gm of water from 14.5 to 15.5°C. Work is generally defined as

$$W = F \times d$$

where F is the force applied on the body and d is its displacement under the application of this force. One joule is defined as the work done in moving one kilogram through one meter in the absence of any external forces. The work done by a system is taken to be positive, while work done on a system is negative. Work and heat are inter-convertible just as, in general, different forms of energy are inter-convertible. Thus one calorie equals 4.186 joules. Work and heat, unlike pressure, volume, mass, etc., are not intrinsic properties of a system.

A system can go from one state to another through an infinite number of different processes or paths. The values of Q and W for each one of these paths are different. However, the difference, $Q—W$ remains the same and is independent of the path taken. Thus

$$dU = dQ - dW = U_f - U_i \tag{1.1}$$

where dU is the change in the internal energy, an intrinsic property of the system. Equation (1.1) is the first law of thermodynamics. It can be paraphrased as 'energy can neither be created nor destroyed'.

The first law tells us about the energy balance but not about the direction in which the process proceeds. The fact that every thermodynamic process always goes in a particular direction may be illustrated as follows. Consider an iron rod lying on a table. It will never happen, spontaneously, that all the heat energy in the rod gathers together at one end making that end of the rod hotter than the other. On the other hand, if there is an imbalance in the heat at the two ends of the rod to begin with,

then it always happens that, as time proceeds, the difference in temperature between the two ends becomes smaller and smaller and the heat at both ends tends to equalize. The above process therefore has a particular direction. This intuitive concept is codified as the second law of thermodynamics, which makes it possible to predict the direction of the natural process. This law is stated in many different ways. One of the conventional statements is "It is not possible to convert heat completely to work with no other change taking place". It is also stated in terms of heat flow, as follows: "It is impossible for heat to flow from one body to another which is at a higher temperature, with no other change taking place".

The third law of thermodynamics states that the entropy of a pure solid or liquid is zero at the absolute zero temperature. It is sometimes stated as follows – "It is impossible to attain absolute zero temperature through a finite number of operations."

There are several thermodynamic variables that are used to describe a thermodynamic system. Some of them are described in detail below.

1.7.1 Entropy

Entropy is a variable used in a precise statement of the second law of thermodynamics. It is defined as a thermodynamic function whose change is independent of the path of transformation of the system and is given by

$$dS = \delta Q/T \tag{1.2}$$

i.e. the change in entropy dS is measured as the ratio of the change in the heat energy δQ to the temperature at which the change takes place in a reversible process. The difference in entropy between any two states can be obtained by integrating the above equation

$$dS = S_f - S_i = dQ/T$$

where the integration is carried out over the reversible path of the process. S_f and S_i are the final and initial states of the system respectively. Thermodynamic equations most often involve the difference in entropy between the initial and final states of the system, rather than the absolute entropy of a system and it is this quantity which is of interest in almost all the cases. This difference is independent of the path taken during the process. Entropy (S), like temperature (T) and internal energy (U), is an intrinsic property of the system and, like T and U, is called a state function describing the state of the system. In the case of a reversible process, the change in entropy is calculated in a straightforward way, using the above equations. In the case of an irreversible process, the change in entropy between two equilibrium states is calculated by finding a reversible path between the two states and calculating the entropy change for that path. The second law of thermodynamics can be restated to say that the entropy of an irreversible process in an isolated system always increases, i.e.

$$\Delta S = S_f - S_i > 0$$

Thus naturally occurring spontaneous irreversible processes, such as biological processes, always result in an increase in entropy. Boltzmann showed that entropy could be defined in such a way that it will be a measure of order in a system

$$S = k \log (w)$$

where k is the Boltzmann's constant and w the number of configurations a system can take. A larger number of configurations imply a less ordered system. Thus to state that a spontaneous process moves toward the maximum entropy is the same as saying that the system will move toward maximum disorder or randomness.

1.7.2 Enthalpy

The enthalpy of a system H is defined as

$$H = U + PV$$

where U is the internal energy, P the pressure and V the volume. P and V are intrinsic properties and hence thermodynamic parameters. Their product is expressed in units of energy. Therefore enthalpy is also expressed in units of energy and is known as the heat content of a system. Since H *is a* combination of state functions and parameters that are independent of the path of transformation of the system, enthalpy too is independent of the path. The change in enthalpy is obtained by considering the differential form of the equation, which is

$$dH = dU + PdV + VdP$$

At constant pressure $VdP = 0$. Therefore,

$$dH = dU + PdV \tag{1.3}$$

But from the first law of thermodynamics we know that

$$dQ = dU + dW$$

where W, the work done, equals PdV (at constant pressure). Therefore

$$dQ = dU + PdV = dH \tag{1.4}$$

Thus enthalpy is an intrinsic property of the system. Its increase is equal to the amount of heat absorbed from the environment at constant pressure. A reaction is exothermic when heat is let out to the surroundings and the enthalpy change is then negative. On the other hand, in an endothermic reaction heat is absorbed from the surroundings, and the enthalpy change is positive.

1.7.3 The free energy of a system

A linear combination of thermodynamic state functions such as internal energy U and entropy S is also a state function. Such combinations are known as thermodynamic potentials. Enthalpy, defined above, is one such potential. The Helmholtz free energy is another, used widely in biology. It is defined as

$$F = U - TS$$

The change in free energy is thus

$$dF = dU - TdS - SdT$$

The physical meaning of this differential equation is that a change in free energy will involve a change in internal energy, in temperature and in entropy. At constant temperature, i.e. for an isothermal process, $dT = 0$. In that case

$$dF = dU - TdS \tag{1.5}$$

Now the first law of thermodynamics states that

$$dU = dQ - dW$$

But, from equation (1.2)

$$dQ = TdS$$

Therefore

$$dU = TdS - dW \tag{1.6}$$

Now combining (1.5) and (1.6) we have

$$-dF = dW \tag{1.7}$$

Therefore the work done under isothermal conditions always leads to a decrease in the Helmholtz free energy (as inferred from the negative sign of *dF).* It is also seen from the above relations that when the system is in an equilibrium state, the entropy is at a maximum and the free energy is at a minimum.

Another thermodynamic potential function that is used to describe biological processes is Gibb's free energy and is defined as

$$G = U - TS + PV$$

Under constant pressure and temperature (i.e., $dT = 0$, $dP = 0$) the change in this potential is given by the appropriate differential form of the equation,

$$dG = dU - TdS + PdV = dH - TdS \tag{1.8}$$

For reversible processes, using equation (1.6), we have

$$-dW_{max} = dU - TdS$$

since the maximum work is done in such processes. Thus

$$dG = -dW_{max} + PdV$$

Therefore

$$-dG = dW_{max} - PdV = dW_{net}$$

At equilibrium *dG, dS* and *dF* are all zero. Once again the decrease in free energy is a measure of the useful work done under conditions of constant pressure and temperature.

1.7.4 Chemical potential

In a closed system the mass and chemical composition of the constituents remain constant. But in an open system, such as a biological process, the mass and chemical composition may change. Hence, thermodynamic quantities like *U, S, F* etc. will also depend on the quantities of chemicals present in the system. Let $n_1, n_2, n_3 \ldots$ be the quantities of the 1st, 2nd, 3rd ... components. Then the change in free energy by the addition of 1 mole of component i to the system at constant pressure and temperature, when the quantities of the other components remain unaltered, is given by

$$\mu_i = \left(-\frac{\partial G}{\partial n_i} \right)_{T,P,n_j,j \neq i}$$

μ_i is known as the chemical potential of the component *i.* Generally $n_1, n_2, n_3 \ldots$ are expressed in gram-moles. For a one component system, the chemical potential is nothing but the Gibb's free energy measured per mole of the component. For a multi-component system the total Gibb's free energy

$$G = \sum_i \mu_i n_i$$

The free energy of a system cannot be defined as an absolute number, since what is measured is only the change in the free energy during a process. Therefore a relative scale is defined for every element in its stable state, such that the free energy is zero at a temperature of 2S°C and a pressure of 1 atmosphere. The free energy of a compound is therefore nothing but the change in free energy due to its formation from its elements. The change in the standard free energy in the course of a reaction (denoted $\Delta G°$) can be computed from the standard free energy of formation of reactants and products.

$$\Delta G° = - RT \log K$$

where K is an equilibrium constant, R the gas constant and T the temperature. For the reaction

$$A + B \text{ (reactants)} = P + Q \text{ (products)}$$

if C_A, C_B, C_P, C_Q are the concentrations of the components, then

$$K = (C_A C_B)/(C_P C_Q)$$

When $\Delta G°$ is negative, i.e. if the concentration of the products is higher than that of the reactants at equilibrium, then the reaction will proceed spontaneously in the direction of the products and is exothermic. If it is positive, it requires energy for the reaction to go in the direction of the products i.e. the reaction is endothermic. Thus standard free energy change is an indicator of the direction of the reaction. However, it does not quantify the rate of reaction.

1.7.5 Oxidation- reduction potential

Oxidation-reduction reactions, which are very common in biochemical processes, involve the transport of electrons. A substance is said to be 'oxidized' when it looses electrons and said to be 'reduced' when it gains electrons. A biological reaction takes place in an aqueous medium and hence the reactant can pick up a proton from the surrounding medium and become reduced. Or it can bind to an oxygen atom and become oxidized. Consider the following reaction.

$$A_{oxid} + B_{red} \leftrightarrow B_{oxid} + A_{red}$$

Since in an oxidation-reduction reaction there is an exchange of electrons, the free energy of the reaction is expressed in units of electric potential. The above equation can be split as follows:

$$A_{oxid} \leftrightarrow A_{red}$$

where A_{oxid} and A_{red} are called a redox couple. Similarly,

$$B_{oxid} \leftrightarrow B_{red}$$

The change in free energy for one-half of the reaction is

$$\Delta G° = \Delta G° + RT \log (A_{oxid}/A_{red}) = \Delta G° + RT \log (\alpha/(1-\alpha))$$

where α is the fraction of A that is oxidised. Converting this to electrochemical potential, we obtain

$$E_A = E°_A - (RT/ZF) \log (\alpha/(1 - \alpha)) \qquad (1.9)$$

where E is the oxidation-reduction (redox) potential, F is Faraday constant, and Z the number of transported electrons in the reaction. The redox potential is measured against a standard, which is generally a hydrogen electrode. The potential E measured while inserting the hydrogen electrode in a 1 N acid solution (i.e., pH = 0) is taken to be zero. Against this standard, the redox potential of other substances can be measured. As biological processes take place at pH = 7 (i.e. neutral pH) it is more

convenient to use the redox potential measured at this value as the standard. For such a measurement equation (1.9) is generally written as

$$E_R = E_m - (RT/ZF) \log (\alpha/(1 - \alpha))$$

where E_R is the redox potential measured against a standardized electrode and E_m the midpoint potential, measured when half of the substance is in the oxidized state and the other half is in the reduced state. A strong reducing agent (such as NADH) has a negative redox potential while a strong oxidizing agent (such as O_2) has a positive redox potential.

1.8 Radioactivity

An atom consists of a nucleus surrounded by electrons. The nucleus in turn consists of protons and neutrons. The atomic number Z of the atom is the number of electrons present, which is also equal to the number of protons as the atom is electrically neutral. The mass number A of the atom is given by the sum of neutrons and protons, i.e. $A = Z + N$, where Z = atomic number and N = number of neutrons. Atoms with the same atomic number but different mass numbers are known as isotopes. The number of isotopes for a given element may vary from two to as many as 20. For example, hydrogen has three isotopes, $^{1}_{1}H$ (hydrogen), $^{2}_{1}H$ (deuterium) and $^{3}_{1}H$ (tritium). The superscript stands for the mass number. In general an element is represented $^{A}_{Z}Z$, where A is the mass number and Z the atomic number. In an atomic nucleus when the protons and neutrons are equal in number, the isotope is generally a stable one. However, elements with higher atomic numbers frequently have a neutron to proton ratio of more than one. These isotopes are unstable and are known as radioisotopes. Radioisotopes are found in nature (e.g. uranium, plutonium, thorium and radium) though they are more often made artificially. Radioactive isotopes emit particles and/or electromagnetic radiation to become a stable isotope. This process is known as radioactive decay. A neutron can emit an electron (also known as a negatron to indicate is nuclear origin) to become a proton

$$\text{neutron} \rightarrow \text{proton} + \text{negatron}$$

Similarly, a proton can become a neutron by emitting a positron. A positron has the same mass as an electron but is positively charged.

$$\text{proton} \rightarrow \text{neutron} + \text{positron}$$

Both the above processes alter the N/Z ratio. For example

$$^{14}_{6}C \rightarrow {}^{14}_{7}C + \beta\,(-ve)$$

$$^{22}_{11}Na \rightarrow {}^{22}_{10}Ne + \beta\,(+ve)$$

Sometimes the term 'beta particle' is used to denote either a positron or a negatron. Radioactive decay can also result in the emission of α particles and γ-rays. An α particle is a helium nucleus (two protons and two neutrons). The emission of an α particle leads to a reduction in mass number by 4 and atomic number by 2, e.g.

$$^{226}_{88}Ra \rightarrow {}^{222}_{86}Rn + {}^{4}_{2}He\,(\alpha\text{-particle}) \rightarrow\rightarrow {}^{206}_{82}Pb$$

As the product of the first decay, radon, is also radioactive, it initiates a further decay series which finally leads to the formation of lead. γ-ray emission does not lead to any change in atomic number or mass. γ-rays are electromagnetic waves at the extreme short wavelength, high frequency end of the

spectrum. **γ-ray** emission is frequently accompanied by **α** and **β** particle emission. **γ-ray** emission arises when the excited nucleus returns to its ground state.

1.8.1 Rate of radioactive decay

Radioactive decay always occurs at a specific rate which is characteristic of the source and follows an exponential law. If we consider a sample containing N nuclei then the decay in a given time will be proportional to the number of nuclei available for decay. The rate of change in the number N of radioactive atoms is given by the differential equation

$$-\delta N/\delta t = \lambda N$$

where λ is the proportionality constant or the decay constant.

$$\lambda = -(\delta N/N)/\delta t$$

λ may also be considered the fractional rate of decay during the short period of time in which N may be treated as a constant. Integrating this equation over time t, we get

$$\log (N_t/N_0) = -\lambda t$$

or

$$N_t = N_0 \exp(-\lambda t)$$

where N_t is the number of radioactive atoms present at time t and N_0 = the number present at time $t = 0$ (i.e. the initial number). From the last equation we see that there will always be some parent nuclei present after any finite time, however long. It is therefore usual to specify the half-life of a radioactive material rather than its lifetime. The half-life of a radioactive nuclide is defined as the time taken for one half quantity of the sample to decay from the parent nuclide to the daughter nuclide (Figure 1.13). Substituting this definition into the above equation we have $N_t = 0.5N_0$. Thus half-life is given by

$$T^{1/2} = \log 2/\lambda$$

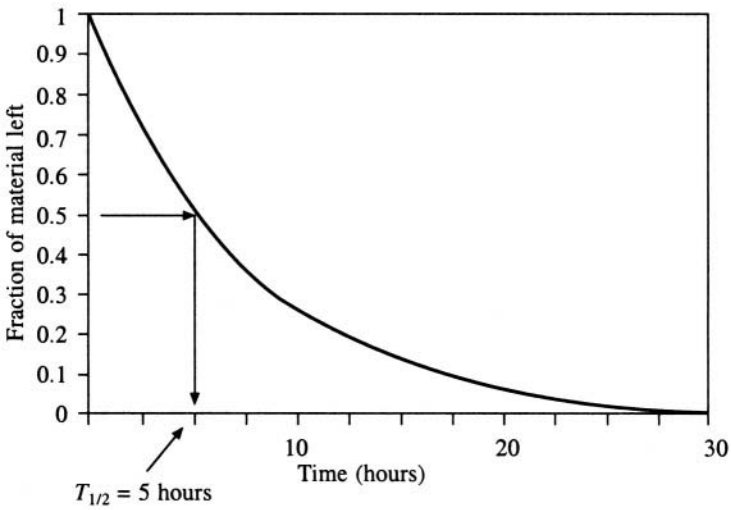


Fig. 1.13 Radioactive decay curve with decay constant λ = 0.1386, corresponding to a half-life of 5 hours.

The half-lives of radioactive materials vary from 10^{19} years to 10^{-7} secs. The half-lives of some of the isotopes are:

^{3}H — 12.26 years ^{32}P — 14.20 days

^{14}C — 5760 years ^{42}K — 12.40 hours

1.8.2 Measurement of radioactivity

Apart from the half-life, radioactive decay can also be expressed as disintegrations per minute (dpm) or per second (dps). The unit of radioactivity used is curie C_i and is defined as that quantity of radioactive substance in which the number of nuclear disintegrations is equal to that in 1 gram of radium. Thus 1 curie = 3.7×10^{10} dps. Millicurie (1 mC_i = 1/1000 of a curie) and microcurie (1 μC_i = 1/1000000 of a curie) are the units generally used in biological studies. Specific activity is the amount of radioactivity per unit mass of the sample. Hence, the unit of specific activity is mC_i/mg or μC_i/mg. The amount of radiation that falls on a sample can be measured in roentgens or in rads. Roentgen is generally used in cases where electromagnetic radiation is involved, while rad can be used both for radiation and particles. One rad is defined as the dose required to produce 100 ergs of absorbed energy per gram of target. A dose of one rad can increase the temperature of one gram of material by 2×10^{-6} K.

Radioactivity can be measured by counting devices or by photographic methods. In both cases advantage is taken of the ability of the radioactive emission to interact with matter. Geiger-Muller counters are based upon the ability of the radioactive emission to ionize matter. Scintillation counters are based on its ability to excite an atom leading to emission of visible light as the atom returns to the ground state. The photographic method is advantageous when we require a permanent record of the experiment conducted. In this method the ionizing radiation from a radioactive source interacts with a photographic emulsion (silver halides) and leaves an image. Autoradiography, which uses photographic method, is a sensitive technique by which the distribution of radioactivity in a biological specimen can be determined, for example, the sites of localization of a radiolabeled drug in the body.

1.8.3 Effects of radioactivity on matter

a particles ($^{4}He^{2+}$), due to their size, are slow and interact with atoms causing ionization and excitation. Ionization takes place when an outer electron of an atom is completely removed. Excitation takes place when the electrons are raised to a higher excited orbital. Though the initial energy of alpha particles is reasonably high (3 to 8 Mev) their energy is dissipated very quickly and hence they are not very penetrating. β particles with a negative charge (negatrons) are small compared to α-particles and hence their interaction with matter is very much less. They are thus more penetrating. They are rapidly moving particles causing ionization and excitation, though less than α-particles. γ-rays are electromagnetic radiation and carry no charge or mass. They travel very large distances without appreciable loss of energy, and are thus highly penetrating. γ-rays also interact with matter and produce secondary electrons. A low energy γ-ray can transfer all its energy to an orbital electron, which is then ejected out of the atom as a photoelectron. This phenomenon is known as photoelectric absorption of the radiation. A medium energy γ-ray transfers only a part of its energy to an orbital electron (which is then ejected) and proceeds further with reduced energy. The change in energy between the impinging photon and the emitted photon can be measured as a change in its wavelength and is known as the Compton effect. A high energy γ-ray, when it interacts with a nucleus, transfers all its energy to it and excites it. The excited nucleus in turn emits another photon of the same or lesser energy. The photon may then spontaneously decay to produce a positron-negatron pair in a process known as 'pair-production'.

1.8.4 Biological effects of radiation

Ionizing radiation can affect both somatic and sex cells. The somatic effects depend on the amount of radiation and it can cause reddening of the skin, loss of hair, etc. It is also known that radiation can cause carcinogenesis. Further, biological specimens consist of a large amount of water, and irradiation of these can lead to hydrolysis

$$H_2O \rightarrow H + OH \quad \text{or} \quad H_2O \rightarrow H_2O^+ + e^-$$

Unlike genetic damage, which is caused by even very small amounts of radioactivity, serious somatic effects occur only with exposure to higher doses. Even so, a dose of about 500 rads can immediately kill a human being. Further, irradiation can release free electrons which in turn can produce secondary electrons by collision. These free electrons can cause a great deal of damage in the cell. Genetic effects occur when the reproductive cells are irradiated, leading to mutations in DNA The damage to DNA due to irradiation can be in the form of breaks in the strands or alterations in the base sequence. For a given dose, mRNA is the most vulnerable, while DNA synthesis itself is the least affected. Thus the biological effects of radioactivity could be disastrous to life.

1.8.5 Applications of radio isotopes

^{3}H, ^{14}C and ^{32}P isotopes are used in autoradiography of biological samples. Radioactive isotopes used in biological experiments are known as radiotracers. The greatest advantage of the radiotracer method is its extreme sensitivity. Metabolites that occur in very low concentrations in the system can be identified and therefore, radiotracer experiments are frequently performed *in vivo.* Apart from their use in tracing metabolic pathways, in pharmacology radioisotopes are used to study drug accumulation, the rate of metabolism of the drug, and also the metabolic products of the drug under consideration. On the analytical side, enzyme substrate binding studies can be carried out using radio labeled substrates. Radioimmunoassay is yet another area where radioisotopes are used. Radioisotopes have other biological applications like sterilization of food, clinical diagnosis, ecological studies etc. Isotopes like ^{60}Co, ^{226}Ra, ^{222}Rn are used in radiation therapy of cancer.

2

Separation Techniques

2.1 Introduction

The identification, purification and separation of the components of a mixture of molecules is of considerable importance in molecular biology and biophysics. There are several techniques available for this purpose and the two most frequently used described in this chapter are: Chromatography and Electrophoresis.

2.2 Chromatography

Chromatographic techniques can be classified into four main categories based on the type of molecular interactions used for separation, viz. ion-exchange chromatography, adsorption chromatography, partition chromatography and molecular exclusion chromatography. The techniques can also be classified in another way, on the basis of the mode of separation and the support system used, as column chromatography, thin layer chromatography and paper chromatography. The particular technique that is used for a given sample depends on its characteristics. Sometimes, several techniques may be used one after the other to obtain extra pure sample.

The basic principle behind all the different types of chromatography is however the same. Chromatographic separation takes advantage of the fact that the sample distributes or partitions itself to different extents in two different, immiscible phases. This property is described by the partition or distribution coefficient K_d. If we consider two immiscible phases A and B,

$$K_d = \text{(Concentration of the sample in phase A)/(Concentration in phase B)}.$$

The effective distribution is then defined as the total amount of the substance present in phase A divided by the total amount of substance present in phase B. The two immiscible phases could be a solid and a liquid, or a gas and a liquid, or a liquid and another liquid. Generally, one of the two phases is a stationary phase and does not move. The other is a mobile phase and moves with respect

to the first. To understand how chromatography is used to separate and purify molecules, consider the column mode, with a stationary solid phase and a mobile liquid phase. To separate a mixture of molecules, a column is packed with the granular stationary phase to a height of a few centimetres and is then surrounded by the liquid phase. Assume that the column diameter is such that **1 cm^3** of liquid will surround 1 cm length of the column material. First consider the behaviour of a solvent containing a single solute. Let **32 μg** of the sample in **1 cm^3** solvent be added to the column. It would occupy position *A* in Figure 2.1. If the effective distribution coefficient as defined above is 1.0, the solute would distribute itself equally between the liquid phase (the solvent) and the solid phase (the support material in the column). Now, if another **1 cm^3** of the solvent (this is also sometimes more properly called the 'eluant', especially when it is particularly used to elute or bring out the sample) is added to the column, half the solute will move from position *A* to position *B,* since the distribution coefficient is 1.0. There will now be **16 μg** of sample each at positions *A* and *B.* If yet another **1 cm^3** of the solvent is added to the column, the situation will be as shown in the third panel of Figure 2.1, with **8 μg** in position *A,* **16 μg** in position *B* and **8 μg** in position *C.* When this process is repeated, the pattern shown in the other panels in Figure 2.1 will be observed. At every step of the equilibration, 50% of the sample will be left bound to the solid phase at each position. After five equilibration steps the sample is present all through the column but has a maximum concentration at the centre of the column. If the solvent contains two solutes with different partition coefficients, at each step the amount in the solid phase will be different for the two. After a number of such steps the peaks of concentration for the two components will be at different positions along the column, thus achieving a separation of the mixture. If the process is allowed to continue sufficiently long, the different components of the mixture will flow out of the column at different times and can be collected separately. This is the basic principle of chromatography. We now consider different applications of this principle.

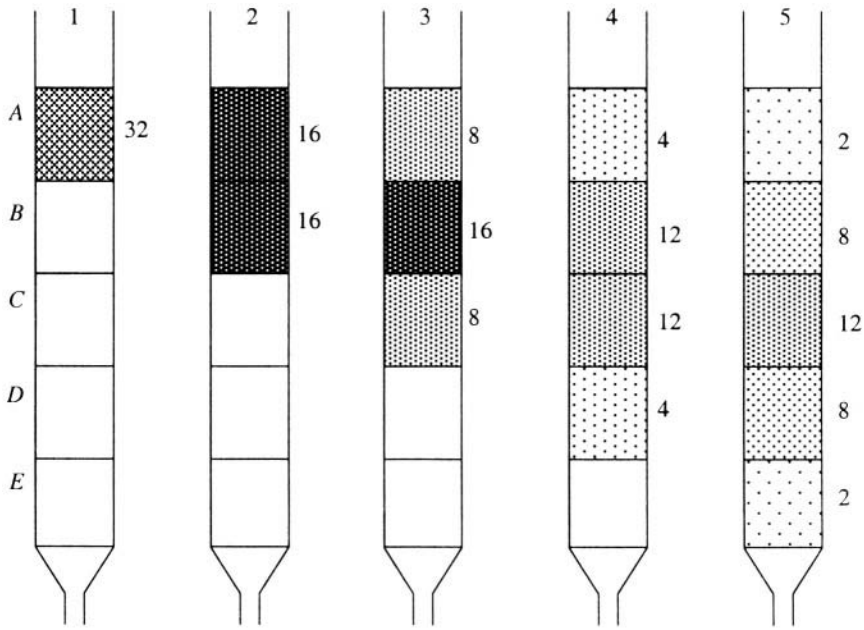


Fig. 2.1 The column chromatographic separation

2.2.1 Column chromatography

Adsorption chromatography, partition chromatography, ion exchange chromatography, exclusion chromatography and affinity chromatography all use the column mode (Figure 2.2), where the stationary phase is packed into glass or metal columns. The nature of the stationary phase depends upon the particular form of chromatography. It is generally an adsorbent, in other words, a resin or a gel.

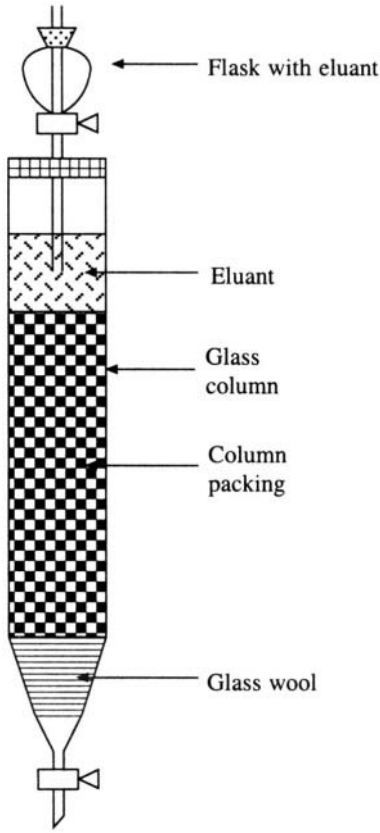


Fig. 2.2 Column mode of chromatography

2.2.2 Thin layer chromatography

In this case, the stationary phase is applied to a glass, plastic or metal foil plate as a uniform, thin layer and the sample is applied at the top edge of the layer using a micro-pipette or syringe. This technique can be used for partition chromatography, adsorption chromatography and exclusion chromatography. The advantage of this method is that a large number of samples can be studied simultaneously. Also, it is quick and simple.

2.2.3 Paper chromatography

In this technique, the cellulose fibres of a sheet of special chromatographic paper act as a support or the stationary phase (Figure 2.3). There are generally two methods that are employed in paper chromatography – the ascending method, and the descending method. In both of them the solvent is placed at the bottom of a jar or tank so that the chamber is saturated with solvent vapour. The chromatographic paper is held vertically inside the tank. In the ascending method, the lower end of the paper dips into the solvent and the sample spot is applied just above the surface of the solvent. As the solvent moves up the sheet of paper by capillary action, the sample constituents move along with it and get separated. In the descending method, the upper end of the paper, where sample is

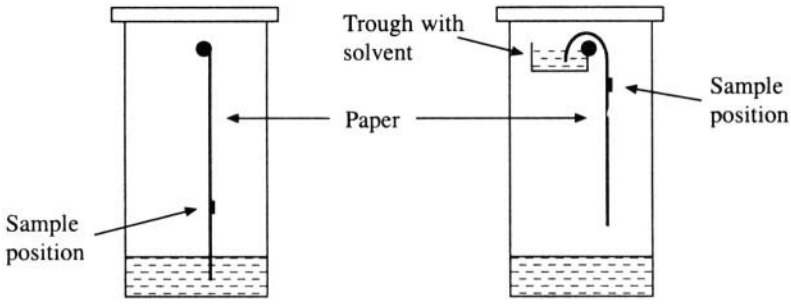


Fig. 2.3 Two types of paper chromatography

applied, is held in a trough of solvent at the top of the tank, and the rest of the paper is allowed to hang down, without touching the solvent at the bottom of the tank. Again by capillary action (and gravity), the solvent moves down the sheet of paper, taking the sample along with it and separating the components. The components are detected by spraying the paper with specific colouring reagents, for example, ninhydrin, which binds only to amino acids and proteins. Sometimes fluorescent dyes are used. For example, ethidium bromide is used with DNA samples. When the paper is subsequently examined under ultraviolet light, the position of the fluorescent or ultraviolet absorbing spots can be seen. The corresponding compounds are identified on the basis of their R_f values where

$$R_f = \text{(Distance moved by solute from the origin/Distance moved by solvent).}$$

The R_f value is the relative displacement of solute with respect to the solvent front and is a constant for a given compound under standard conditions. Thus by looking up a table of R_f values, it is possible to identify the compound after separation.

As mentioned earlier, chromatographic techniques are also classified on the basis of the molecular property of the support system or sample that is used in the separation. Some of these are considered in greater detail below.

2.2.4 Adsorption chromatography

This technique separates the sample into its components based on the degree of their adsorption[2] by the adsorbent and on their solubility in the solvent used. Silicic acid (silica gel), aluminum oxide, calcium carbonate and cellulose may be used as the stationary phase in this method. The optimum combination of adsorbent and eluting solvent will depend upon the type of sample separation under consideration. For example, hydroxyapatite (calcium phosphate) is used for DNA separation as this adsorbent can bind double stranded DNA but not single stranded DNA or RNA. Normally any organic solvent can be chosen as the mobile phase, e.g. hexane, heptane, proponal, butanol etc.

2.2.5 Partition chromatography

There are two types of partition chromatography. (a)Liquid-liquid chromatography: In this type of partition chromatography, the material separation is achieved based on its differing solubilities in two (immiscible) liquid phases. The sample is dissolved in a relatively non-polar solvent and is passed through a column that contains immobilised water or some other polar solvent in a supporting matrix such as silica gel. When the sample moves through the column it gets partitioned between the mobile phase and the immobile, polar phase. In the reverse phase liquid-liquid chromatographic method, the stationary phase is a non-polar substance like liquid paraffin supported by the matrix. High Performance Liquid Chromatography (HPLC) is a liquid-liquid partition chromatographic technique. It is widely used for the separation of biological compounds either as a normal phase method, or as a reverse phase method. In the latter, it is useful for separating polar compounds. (b) Counter current chromatography: This is also partition chromatography with this major difference, that there is no supporting matrix. The sample is equilibrated between two immiscible solvents (e.g. ethylacetate and water) in a test tube. The partition coefficient is defined as

$$K = \text{(Concentration } C_1 \text{ in the upper phase/Concentration } C_2 \text{ in the lower phase)}$$

After equilibration the upper phase is withdrawn and transferred to a fresh volume of lower phase

[2]Adsorption is the process by which a substance is absorbed on to the surface of another. There is almost no penetration beyond a few molecular layers.

solvent in another tube. The original lower phase is added to a fresh volume of upper phase solvent, and so on. The process is repeated several times, yielding several test tubes, half of them finally containing the upper phase solvent, and half containing the lower phase solvent. A solute which is more soluble in upper phase ($K > 1$) will concentrate in tubes containing the upper phase solvent. If the solubility is greater for lower phase ($K < 1$) then it will concentrate in tubes of the lower phase solvent. Counter current chromatography is successfully used for cell organelle fractionation.

2.2.6 Gas liquid chromatography (GLC)

The two phases used are a liquid and a gas. The method is very sensitive and reproducible. The partition coefficient of a volatile sample between the liquid and the gas phases, as it is carried through the column by the carrier gas, may be largely different from unity. This feature is exploited to separate or purify the compound. The stationary or immobile phase is generally nonvolatile and thermally stable at the temperature of the experiment. Usually organic compounds with high boiling point are coated on a base to form the stationary phase. The base is generally made up of celite (diatomaceous silica). The sample to be separated is packed in a narrow, coiled, steel or glass tube and placed in an oven at an elevated temperature. An inert gas like argon, nitrogen or helium is passed through the tube. The sample is volatilised and is carried away by the mobile gas phase. It is passed over the solid phase packed into another column, and as it does so, the sample distributes itself between the gas and solid phases.

2.2.7 Ion exchange chromatography

Many biological samples like amino acids and proteins have ionisable groups. The samples can therefore be separated based on the net charge acquired by them at a certain pH. The ion exchanger is a resin and can be either a cationic exchanger or an anionic exchanger. The cationic exchanger has negatively charged groups and attracts positively charged molecules. The anionic exchanger has positively charged groups and attracts negatively charged molecules. Commercially available ion exchangers are made by co-polymerizing styrene with divinyl benzene. The two materials become cross-linked, with the permeability of the resin varying with the degree of cross-linking. When this insoluble, cross-linked resin is placed in water, a network of charged groups is formed. When the sample, suspended in the liquid phase, comes in contact with the resin, the charged groups in the sample replace the charged ions on the surface of the resin and the sample becomes strongly bound to the resin. As an example, suppose that the sodium salt of a strongly acidic resin, which contains SO_3^- groups, is placed in a suspension. Normally the Na^+ ions would be bound to the negative sulphate groups in the resin. However, on exposure to the sample, the positively charged groups present in the sample may displace some of the Na^+ ions from the surface. This exchange of ions at the exchange site leads to the distribution of the sample between the solid and liquid phases. The higher the charge on the molecule to be exchanged, the tighter it binds to the resin. At low pH and salt concentration most amino acids will firmly bind to the resin by the following reaction

$$RSO_3^- \ldots Na^+ + N^+H_3R' \leftrightarrow RSO_3^- \ldots N^+H_3R + Na^+$$

The last step in the separation procedure is to elute the bound molecule from the resin by reducing the attractive force between the resin and the charged molecule. This is achieved by changing the pH or the ionic concentration, or by affinity elution. In affinity elution, an ion with greater affinity than the bound molecule is introduced into the system. If more than one type of molecule is bound to the column, a stepwise elution can also be done by using a second solvent, a third solvent and so on. A widely used matrix medium for ion exchange chromatography is DEAE-cellulose column in which

the Diethyl aminoethyl group acts as an exchanger. Cellulose fibres, polyacrylamide gel, crosslinked dextran (also known as sephadex) are other common materials used as the matrix.

2.2.8 Molecular exclusion chromatography

The separation of samples based on their molecular weights is known as molecular exclusion chromatography (MEC), gel filtration or molecular sieve chromatography. The matrix for MEC is made up of polyacrylamide, sephadex or agarose gel. Polyacrylamide and sephadex are supplied in the dehydrated form. When soaked in the solvent, they take it up rapidly and swell up forming a slurry. When this slurry is used for packing a column, pores of uniform size are formed by the gel and are used for separating the sample according to their molecular size and shape (Figure 2.4). A column of the gel (the porous material) is kept in equilibrium with a suitable solvent in which the molecule to be separated is dissolved. Large molecules in this solution, are excluded from the pores, and pass through the interstitial space reaching the bottom of the column quickly. Smaller particles penetrate through the pores of the gel and are distributed in the solvent inside as well as outside the molecular sieve. Smaller molecules are thus retarded and reach the bottom at slower rate. For a given sample the distribution coefficient K_d is dependent upon its size. If the molecule is completely excluded by the sieve then $K_d = 0$, whereas if it penetrates the gel and has accessibility to the inner solvent then $K_d = 1$. For all other intermediate sizes the K_d value will lie in the range 0 to 1. The chromatography column consists of loosely packed resin particles. V_T is the total volume of the column. It is made up of three terms: (1) The void volume V_0 or the volume of the exterior solvent. (2) V_g or the solid volume of the gel particles. (3) V_i or the internal volume of the pores that are accessible to solvent. Therefore

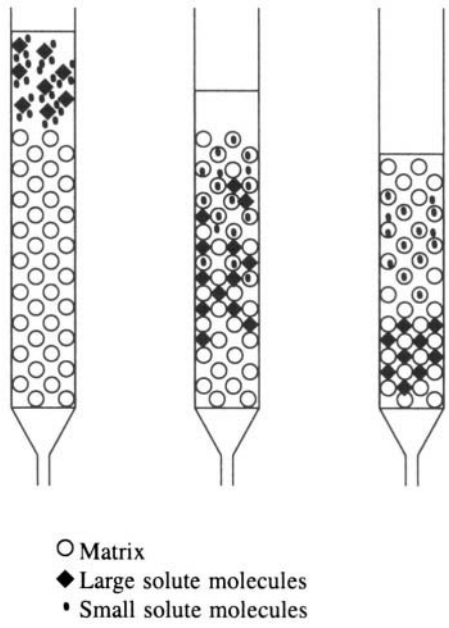


Fig. 2.4 Molecular exclusion chromatography. The column is filled with solvent containing large and small solute particles.

$$V_T = V_0 + V_g + V_i$$

The eluting volume is the volume of solvent that must flow through the column before a particular solute emerges. This is given by the equation

$$V_e = V_0 + K_d V_i \tag{2.1}$$

A solute that is completely excluded will have $K_d = 0$. Therefore $V_e = V_0$ for such a sample, i.e. the eluting volume is equivalent to the void volume. But a solute that can enter the pores will have to displace an additional volume given by $K_d V_i$. Thus the eluting volume in this case is given by equation (2.1). The partition coefficient K_d can be determined from this equation if the elution volume V_e is measured. V_0 and V_i are determined prior to the start of the experiment and are constant for a given column. V_0 is determined by measuring the elution volume of a particle larger than gel pores. V_i is

determined by the weight of water intake by dry resin. Two substances with different K_d values (K_d and K_d') and size can be separated if the sample volume is less than or equal to V_s, where

$$V_s = (K_d - K_d')\ V_i$$

Figure 2.5 shows the relationship between the elution volume and the molecular weight of various proteins when they pass through a sephadex G-200 column. Since the elution volume is related to the partition function by equation (2.1) above, it is possible to correlate K_d to the molecular weight. This relationship is given by the function

$$K_d = -A \log (M) + B$$

where A and B are constants and M the molecular weight. Thus by measuring the elution volume, it is possible to determine the partition coefficient and thence the molecular weight, to finally achieve a separation of the molecules based on the molecular weight.

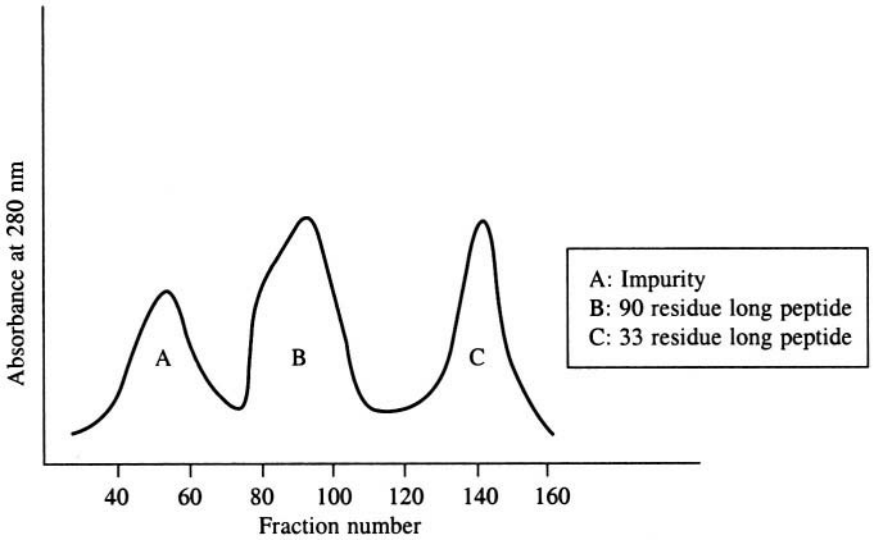


Fig. 2.5 Separation of a mixture of peptides by a sephadex column. Schematic sketch of the absorbance v/s fraction number indicating the difference in movement of the peptides through the column corresponding to the different molecular weights.

2.2.9 *Affinity chromatography*

Separation by affinity chromatography is based on a biological property of macromolecules rather than on a physical property. It is a highly sensitive technique and, in principle, can give absolutely pure sample in a single step. The technique requires that the macromolecule be able to bind to a specific ligand attached to an insoluble matrix under certain conditions, and then detach itself under certain other conditions. In other words the binding should be specific and reversible.

$$\text{Macromolecule} + \text{Ligand} \leftrightarrow \text{Macromolecule} \cdot \text{Ligand}$$

When a complex mixture of proteins is passed through a chromatography column to which a ligand is attached, then the protein with specificity for that ligand alone will bind to the column and the rest of the material will simply flow through. The bound molecule can then be recovered by a change in

the conditions such as pH of the solution, salts or temperature, or by adding a compound which interacts even more strongly to the ligand on the matrix, thus displacing the molecule of interest. The method therefore requires a preliminary knowledge of the biological specificity of the compound to be separated. In practice, agarose, cellulose, and polyacrylamide gels are used as the insoluble matrix. It is generally possible to select a ligand that has absolute specificity for the compound of interest or to select a ligand that has group selectivity. Group selectivity means that the ligand will selectively bind to a closely related group of compounds. For example, 5′AMP can reversibly bind to many NAD dependent dehydrogenases. The other end of the ligand, which is bound to the matrix, should be such that it will not interfere with the reversible binding of the macromolecule. –NH , –COOH and –SH groups are the most common groups used for binding the ligand to the matrix. In addition, there is generally a spacer molecule between the ligand and the matrix, to minimize the interference of the groups in the matrix support material (Figure 2.6). The spacer arm usually has about six to ten carbon atoms and can be hydrophobic (e.g. 1,6-diamino hexane) or hydrophilic (e.g. 6-amino hexanoic acid). Some of the ligands commonly used in affinity chromatography are: 5'AMP (with a specificity for NAD dependent dehydrogenases), lectins (specificity for cells and macromolecules containing N-acetyl-α-galactosamine residues), lysine (specific to RNA) and concanavalin A (specific to glycoproteins and glycopeptides). A highly specific ligand for any macromolecule may be produced by raising antibodies against it and then binding the antibodies to the matrix.

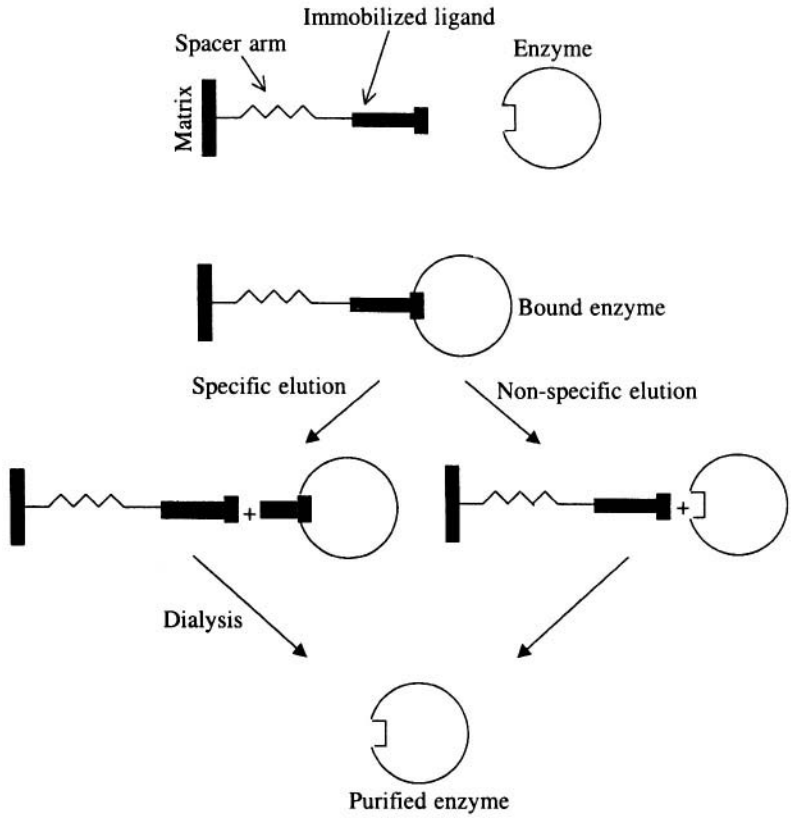


Fig. 2.6 Affinity chromatography. The enzyme attaches itself to a specific ligand that is bound to the matrix. It is then eluted out either specifically or non-specifically.

2.3 Electrophoresis

Important biological molecules like peptides, proteins, carbohydrates and nucleic acids have ionisable chemical groups and hence, under suitable conditions, exist in solution as charged species. Even non-polar molecules can be made to exist as weakly charged species by preparing phosphate, borate or similar derivatives. Molecules with similar charge may have different charge/mass ratios since their molecular weights could be different. Molecules with differing charges or differing charge/mass ratios would migrate at different rates in an electric field. The separation of molecules by utilizing these differences in rates of migration is known as electrophoresis.

If a molecule has a net charge q, the application of an electric field E will result in a force

$$F = qE$$

This force will accelerate the particle in a fluid till a steady state is reached when the frictional force is equal and opposite to the applied force. If v is the steady state velocity then

$$fv = qE$$

where f is the frictional force. Therefore

$$v = qE/f$$

For a spherical molecule with radius a and charge Ze (where e is the charge on electron and Z the atomic number),

$$v = ZeE/(6\pi\eta a)$$

since $q = Ze$ and $f = 6\pi\eta a$, where η is the viscosity of the solution. The electrophoretic mobility u is defined as the velocity per unit electric field.

$$u = v/E$$

However, due to the presence of counter ions in any aqueous solution, this equation does not completely account for the electrostatic mobility. These ions can neutralize some of the charges on the macromolecule if they bind tightly to it. Alternatively unbound or loosely bound ions can modify the ionic strength of the solution and thus alter the electric field that is felt by the polymer. In such a situation the electrophoretic mobility for a spherical particle is modified by a factor K, where K is a function of the screening parameter which has dimensions of L^{-1} (inverse of length). Therefore

$$u = (ZeK)/(6\pi\eta a)$$

Thus mobility will increase with increasing charge on the macromolecule.

Electrophoresis procedures can be generally divided into three categories: (1) Moving boundary electrophoresis. (2) Zone electrophoresis. (3) Continuous flow electrophoresis.

2.3.1 Moving boundary electrophoresis

In this method, the sample is dissolved in a buffer and placed in a cell. An additional quantity of the buffer is layered on top of it. When the electric field is applied, closely related molecules will tend to move as a band since their charges are similar, and boundaries are formed between groups of molecules with different electrophoretic mobility. The separations can be seen by direct optical observations. This method suffers from many experimental difficulties such as errors due to convection. Hence it is not very useful as a preparative technique.

2.3.2 Zone electrophoresis

In this method the mixture to be analysed is applied as a small spot or very thin band to the supporting medium. When electric field is applied the molecules move as distinct bands or zones. The molecules can be identified by staining, *UV* absorption etc. There are several different types of zone electrophoresis. We treat some of them below.

2.3.3 Low voltage electrophoresis

The electrophoresis unit (Figure 2.7) consists of a power pack, which supplies DC current across two electrodes, buffer reservoirs, a support, and a transparent insulating cover. The low voltage power pack can either give a constant voltage or a constant current and has an output of 0 to 500 V and 0 to 150 mA. Either stainless steel or platinum electrodes are used. The supporting medium, which can be paper or cellulose acetate, must first be saturated with buffer and held on a sheet of insulating material like Perspex. The sample is then applied as a small spot using a micropipette. The location of the spot depends on the nature of the molecular mixture present in the sample. For example, if there are molecules with opposite charges, then they will separate and move towards opposite electrodes and hence the sample must be applied at the centre. The two reservoirs on either side of the medium are isolated from the electrodes so that any pH change there does not affect the buffer. After application of the sample, power is switched on for the required amount of time. At the end of the experiment, the migration of the different constituents of the mixture may be analysed. The experiment may be performed in different ways. For example, instead of paper or cellulose acetate, thin layers of silica, alumina or cellulose may be applied on to a glass plate and used as the supporting medium. This technique is similar to TLC (thin layer chromatography) and is known as thin layer electrophoresis (TLE).

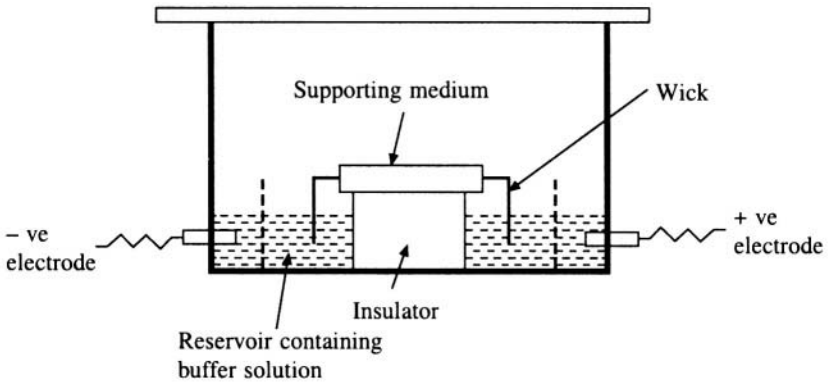


Figure 2.7 Low voltage gel electrophoresis

2.3.4 High voltage electrophoresis

A disadvantage with low voltage paper electrophoresis is that diffusion processes may influence the migration of the molecules so that it is not strictly according to the charges on them. One way to overcome this is to resort to high voltages, of the order of 10,000 volts. Such a voltage can produce potential gradients up to 200 volts/cm. If the separation between the electrodes is d metres and if V is the potential difference in volts then the potential gradient is V/d volts/metre. The force on the charged molecule is then Vq/d newtons, where q is the charge in coulombs. The migration of the molecules is caused by this force and is proportional to it, and an increase in the potential gradient leads to an increased rate of migration. The distance migrated by the ions will depend on the voltage

as well as the amount of time for which it is applied. This technique results in higher resolution and faster separation. But because of the high voltage, it generates considerable heat and the apparatus requires cooling.

2.3.5 Gel electrophoresis

When applied to the separation of a mixture of proteins, simple zone electrophoresis has limited resolving power, due to the similarity in mobility of the components. When zone electrophoresis is combined with molecular exclusion effects using gels, much higher resolutions can be achieved. The molecular sieving property of the gel helps in separating molecules that have similar charge properties but differ in size and shape. Starch, agar, and polyacrylamide gels are used for this purpose. The gels can be run as horizontal slabs or as vertical slabs. Vertical slabs of the type shown in Figure 2.8 are most frequently used. Gels are prepared in glass or Perspex containers. In the case of slab gels, the gel is cast between two clean glass plates that are clamped together but held apart by spacers. The samples to be separated are applied using a micro syringe into the wells set in the gel. In the vertical slab method the two ends of the plate containing the gel are kept in contact with the lower and upper reservoirs of buffer respectively, thus completing the electrical circuit between lower and upper electrodes. The components of the mixture migrate with different mobility on application of an electric field, and can be identified by staining techniques, at the end of the experiment. Normally, a known sample is run along with unknown samples in order to compare the migration characteristics. A major advantage of this method is that even very small quantities of proteins can be separated and identified.

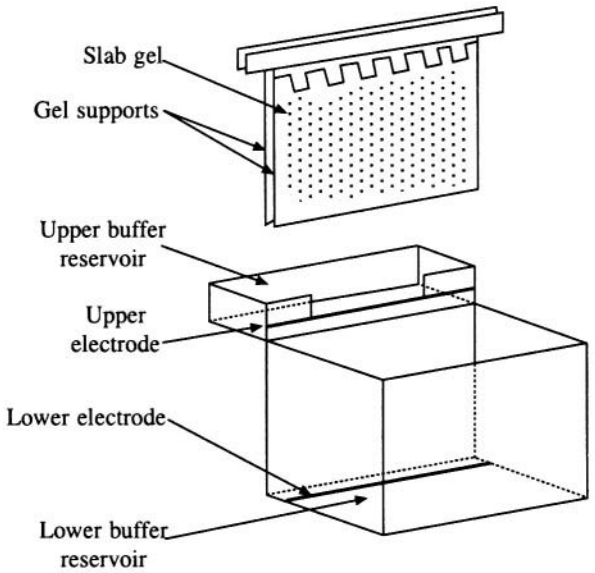


Figure 2.8 Vertical slab gel electrophoresis apparatus. Most of the equipment (except obvious parts such as electrodes and clamps) is made of perspex.

2.3.6 Sodium dodecyl sulphate poly acrylamide gel electrophoresis (SDS-PAGE)

In this technique, all the proteins in a mixture are converted into structures having similar charge characteristics, though they may differ in molecular weight. Sodium dodecyl sulphate is an anionic detergent that binds tightly to proteins and denatures them. When an excess of SDS is added to the

solution (about 1.4 gm per gm of amino acid), then the protein molecules in the mixture are completely surrounded by SDS. The total negative charge acquired due to SDS binding overwhelms the variations in charge in the different protein molecules. Hence, when a drop of the solution is placed on the polyacrylamaide electrophoresis gel, all the protein molecules move towards the anode, with their mobility varying according to their size and shape. Reducing agents like **β-mercaptoethanol** may also be added to break any intrachain or interchain disulphide bonds. The apparent mobility $u(c)$ of the molecule at the gel concentration c is given by

$$\log (u(c)) = -k_x c + \log (u(0))$$

where $u(0)$ is a constant for a set of similar proteins, and k_x is a constant depending on the extent of crosslinking of the gel. It has been shown that the mobility u and molecular weight M are related by the following equation.

$$u = b - a \log (M)$$

where b is a constant which contains the term $u(0)$, and a is a function of gel concentration c. This equation fits the experimental observations extremely well and is used to determine the molecular weight of an unknown sample by running it on a gel side-by-side with another mixture containing standard proteins of known molecular weights.

2.3.7 Isoelectric focussing

Isoelectric focussing is an electrophoretic method where a pH gradient is used along with the voltage gradient. It is extremely useful in separating proteins according to their isoelectric points. The isoelectric point of a protein is defined as the pH at which the average net charge on the molecule is zero. In the electrophoretic gel the region around the anode is kept at a lower pH as compared to the region near the cathode. A pH gradient is established such that the samples to be separated have their isoelectric points within this pH range. At a pH below their isoelectric point, the particles will move towards the cathode, and at a pH above the isoelectric point they move towards the anode. At the isoelectric pH, the particles will remain immobile. Thus, when placed in the combined pH and voltage gradient on the gel, the particles move towards their isoelectric points no matter where they are initially applied. This method gives very high resolution and is especially suitable for separating isoenzymes. Even differences in pH of the order of 0.01 are enough to separate the molecules. Substances known as ampholytes generate the pH gradient between the electrodes. Ampholytes are molecules with both postive and negative charges, for e.g. polymers containing several amino and carboxyl groups. Ampholine, Biolyte and pharmalyte are some of the commercially available ampholytes. When an electric field is applied to a mixture of ampholytes with a wide range of isoelectric points in a column or gel, they distribute themselves according to their charge and thus establish a pH gradient. A small amount of sample may then be added to the medium. The different proteins in it migrate along the column or gel till they reach the pH corresponding to their respective isoelectric points. Each protein then remains at that position as a sharp band and may be viewed and analysed by staining, absorbance etc.

2.3.8 Continuous flow electrophoresis

In continuous flow electrophoresis the experiment is conducted in free solution without a supporting medium. This reduces frictional effects between the sample ions and the solution and causes rapid migration of the components of the mixture. This is a form of the moving boundary method. The sample, in the carrier buffer, is applied through the central tube as a continuous flow. The electric

field is applied between the outer and inner concentric cylinders and the outer cylinder is rotated to maintain a stable flow of the buffer solution. Electrophoresis takes place continuously as the sample is carried upwards by the flow of carrier buffer from the bottom of the inner tube. As it moves upwards the components are separated radially. At the top a series of radial slits separates the buffer stream into thirty or more fractions, each containing a separated portion of the original mixture. Large-scale separations can be achieved by this method. This technique is therefore used for enzyme purification on an industrial scale, blood plasma fractionation. In the laboratory it is used for collection of large amounts of proteins on a preparative scale.

3

Physico-Chemical Techniques to Study Biomolecules

3.1 Introduction

Biological macromolecules have molecular weights of the order of 10,000 daltons. Their functions depend on their three dimensional structures. It is therefore useful to know the structure, size and shape of these molecules in order to understand the mechanism of function. There are several techniques that can give information on these aspects. Table 3.1 lists some of the physical techniques that are used in studying large biomolecules.

Table 3.1 Physical techniques used to study biological molecules and the information they yield

Technique	Information obtained
1. Osmotic pressure	Molecular weight
2. Diffusion	Molecular volume
3. Sedimentation	Molecular weight and volume
4. Viscosity	Molecular shape
5. Birefringence	Molecular shape
6. Optical Rotatory Dispersion	Helical content
7. Light scattering	Molecular size and weight
8. X-ray scattering	Molecular size and shape
9. X-ray crystallography	Three dimensional structure (atomic positions)
10. Electron Microscopy	Size and shap
11. Electromagnetic absorption, nuclear magnetic resonance, electron spin resonance	Intramolecular forces, Orientation of groups, Three dimensional structure

In this chapter we will focus on some of the hydrodynamical methods listed above and discuss in detail the principles underlying them.

3.2 Hydration of Macromolecules

The molecular volume V of any isolated molecule is given by

$$V = (M \langle v \rangle_0)/N_0,$$

where M is the molecular weight, and $\langle v \rangle_0$ is the specific volume, equal to $1/\rho$, where ρ is the density. N_0 is Avagadro's number. However, in the case of proteins and nucleic acids, there may be counter ions and water molecules tightly bound to the molecule and therefore it cannot be treated in isolation. It is normal laboratory practice to store and use proteins and nucleic acids in ~ 0.1 M buffered salt solutions, thereby simulating approximately the physiological environment. Hence, when we study the thermodynamic properties of a macromolecule solution we must treat it as a three component system consisting of salt, water and the macromolecule. Such a multi-component system is very difficult to deal with. However, since we are mainly interested in the interaction of the macromolecules with water, we may approximate the system as consisting of two components: water arid macromolecule. The total volume of such an ideal solution is

$$V_{tot} = g_1\langle v \rangle_1 + g_2\langle v \rangle_2$$

where g_1 and g_2 are the weights of water and macromolecule respectively in grams, and $\langle v \rangle_1$ and $\langle v \rangle_2$ are the specific volumes. In addition to the bound water molecules, we may have bulk water, which may occupy holes or cavities in the macromolecule, thus changing its shape and size. Taking all this into consideration, the volume occupied by the hydrated macromolecule is given by

$$V_h = (M/N_0)(\langle v \rangle_2 + \delta_1\langle v \rangle_1)$$

where (M/N_0) gives the weight of a single macromolecule. $\langle v \rangle_2$ is now the partial specific volume of the solute and is defined as the change in volume of the solution when a small amount of solute is added at infinite dilution, $\langle v \rangle_1$ is the partial specific volume of water $= 1/\rho$, and δ_1 is the hydration in grams of bound water per gram of macromolecule. The hydration value of a protein gives an estimate of the amount of water that is bound to a protein. A value of the order of 0.3 to 0.4 g of water per gram of protein is usually assumed, when discussing the hydrodynamic properties of most macromolecules

3.3 Role of Friction

Hydrodynamic properties of macromolecules like diffusion, viscosity and sedimentation are affected by the frictional forces between molecules of the protein and those of the solvent. Since this frictional force is in opposition to motion we can include this in the equation of motion as

$$F - fv = m(dv/dt)$$

where f is the frictional coefficient, (dv/dt) is the acceleration and m is the mass of the molecule[3]. In the case of spherical particles, the translational frictional force f is proportional to the fluid viscosity η and radius r of the particle. Thus the coefficient of friction for spherical particles, known as Stokes law, is

[3]This is the general equation of motion $F = m(dv/dt)$ or (mass × acceleration) modified to take into account the opposing frictional force which is taken to be proportional to the velocity v.

$$f_{sph} = 6\pi\eta r$$

This relation is true when the interaction of the particle with the fluid is strong. In the other case, when there is virtually no interaction at all between the fluid and the particles, the friction coefficient is

$$f_{sph} = 4\pi\eta r$$

The frictional force can also act as a damping force for the rotational motion of the particle and in that case the coefficient is given by

$$f_{rot} = 6\eta V$$

where V is the volume. Since most biological molecules are not spheres, the previous equations will have to be modified for the shape factor. As a significant fraction of macromolecules are globular, compact irregular bodies or an ellipsoid may be a better approximation than a sphere. Ellipsoids are classified into prolate and oblate ellipsoids. An oblate ellipsoid is disc shaped and is obtained by rotating an ellipse about its minor axis (Figure 3.1). A prolate ellipsoid can be taken as an approximation for a rod and is obtained by rotating an ellipse about its long or major axis. For equal volumes, ellipsoids have greater surface area compared to spheres and hence their frictional coefficients will be larger.

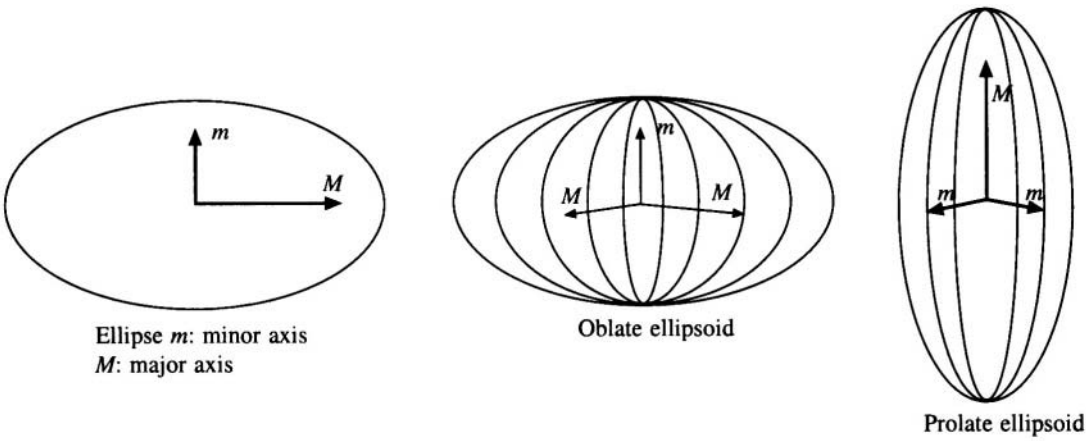


Fig. 3.1 Oblate and prolate ellipsoids

3.4 Diffusion

The frictional coefficient f comes into effect when a molecule moves through a medium. The movement of the molecule could be either diffusion or sedimentation and the driving force can be the concentration gradient, the force of gravity or the centrifugal force. Diffusion can be considered as movement of molecules from a higher concentration to a lower concentration (Figure 3.2). According to Fick's law, the rate of diffusion across a boundary (dn/dt: the number of molecules which pass through a cross section A in unit time) for a single solute component diffusing in a system at constant temperature and pressure is given by

$$dn/dt = -DA(\delta C/\delta r)$$

where $\delta C/\delta r$ is the concentration gradient and D is the diffusion coefficient. A concentration gradient

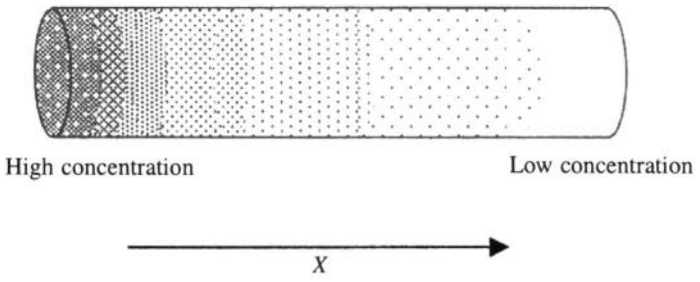


Fig. 3.2 Diffusion of molecules from higher concentration to lower concentration.

implies that the concentration of the molecules (i.e. the solute) varies with distance r in one dimension. The diffusion coefficient D is given by

$$D = kT/f \tag{3.1}$$

where k is the Boltzmann's constant, T the absolute temperature and f the frictional coefficient. Thus D is dependent on f. The concentration of the solute or the concentration gradient can be measured as a function of position, by the absorbance at various points. Alternatively it can be measured as the refractive index gradient using a light scattering technique known as Schlieren optics. Fick's second law of diffusion gives the concentration gradient across the boundary as

$$(\delta C/\delta r)_t = C_0/(4\pi Dt)^{1/2}\exp(-r^2/4Dt) \tag{3.2}$$

where C_0 is the total solute concentration difference across the boundary. In the graph of concentration gradient versus distance, plotted at different times (Figure 3.3), the downward change in the concentration

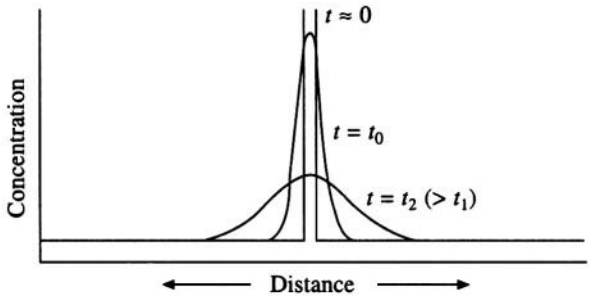


Fig. 3.3 Plot of the concentration as a function of distance at three different times. At $t = 0$, the sample is applied at the centre and is concentrated only there. It starts diffusing until at $t = t_2$ it is spread out over a wide area.

gradient as time passes is clearly observed. The x coordinate of this graph is defined in such a way that the maximum value H_m of the gradient is at $x = 0$. The rate of change of concentration gradient with time ($d/dt\ (\delta C/\delta r)$) is zero at the maximum value of the gradient. Differentiating equation 3.2 with respect to time and setting it equal to zero, we obtain

$$H_m = C_0/(4\pi Dt)^{1/2}$$

When H_m^2 is plotted against $1/t$, the slope of the straight line so obtained will give the value of D, which in turn can be used to determine the molecular weight using equation (3.3), provided the partial

specific volume $\langle v \rangle$ is known. $\langle v \rangle$ is defined as the volume occupied by 1 g of the solute in the solution. For non-hydrated spheres the volume V is given by

$$V = (M \langle v \rangle)/N_0$$

where M is the molecular weight and N_0 is Avagadro's number. But $V = (4/3)\pi r^3$. Therefore

$$\begin{aligned} M &= (4\pi r^3 N_0)/(3\langle v \rangle) \\ &= (4\pi N_0/3\langle v \rangle) \times (f/6\pi\eta)^3 \\ &= (4\pi N_0/3\langle v \rangle) \times (KT/6D\pi\eta)^3 \end{aligned} \tag{3.3}$$

The last step follows from Stoke's law, where η is the viscosity. Thus an experimental measurement of the diffusion constant of a molecule enables the calculation of its molecular weight. Diffusion experiments are not easy to perform, due to experimental difficulties such as the slow rate of diffusion of macromolecules (diffusion constants are of the order of only 10^{-7} to 10^{-6} cm^2/sec), gravitational instabilities, convective mixing, etc.

3.5 Sedimentation

The sedimentation of a suspension of particles in a liquid can be achieved either by the force of gravity or by an applied centrifugal force. For example, if a muddy solution is left to stand undisturbed, it will separate out after some time as a top layer of water and a bottom layer of mud. This sedimentation of mud particles is due to the force of gravity. However, when we perform a sedimentation experiment with macromolecules, the force of gravity is insufficient to make the molecules sediment. Under these circumstances the solution containing the macromolecules is subjected to centrifugation in an ultracentrifuge. An ultracentrifuge can rotate at speeds of several thousand revolutions per minute leading to a centrifugal force of the order of 3,00,000 times the acceleration due to earth's gravity or 3,00,000 g. The force on each particle is given by Newton's second law

$$F = ma$$

where m is the mass of the particle and a is its linear acceleration. As the particles are spun in an ultracentrifuge, the radial or centrifugal force F_C is given by

$$F_C = m\omega^2 r$$

where ω is the angular velocity, and r is the radial distance of the particle from the axis of rotation. As the particles sediment, diffusion effects also take place simultaneously and hence the motion of macromolecules in an ultracentrifuge is the resultant of both the effects. Therefore the mass term in the above equation is replaced by the effective mass which is given by $(m - V\rho)$ i.e. the difference between the weight of the particle and the bouyant force due to displacement of the medium. V is the volume and ρ is the density of the solvent. $V\rho$ represents the mass of solvent displaced by the particle of mass m. Making these substitutions, the centrifugal or radial force is given by

$$F_C = (m - V\rho)\omega^2 r$$

This force is opposed by the frictional force F_r which is proportional to the velocity of the particle

$$F_r = f \times (\text{radial velocity}) = f\,(dr/dt)$$

where f is the frictional coefficient. When the centrifugal force is balanced by the frictional force then

$F_C = F_r$ and the particles will sediment with uniform velocity, since there is no net force on them and thus no acceleration. Writing this condition in another way

$$(m - V\rho)\omega^2 r = f\,(dr/dt) \tag{3.4}$$

in which ω is expressed in radians per second. Multiplying both sides by N_0 (Avagadro's number) we get

$$N_0(m - V\rho) = (N_0 f(dr/dt))/(\omega^2 r)$$

The velocity of sedimentation depends on the size and shape of the sedimenting particle and is proportional to the applied force. Therefore

$$v = s\omega^2 r$$

where s is a proportionality constant depending on shape and size of the molecule and the nature of the liquid. It is known as the sedimentation coefficient. It is defined as the velocity per unit field

$$s = \text{velocity}/\omega^2 r = (dr/dt)/(\omega^2 r)$$

Therefore

$$N_0(m - V\rho) = N_0 fs$$

or

$$N_0 m(1 - V\rho/m) = N_0 fs$$

But $N_0 m$ is the molecular weight M. Therefore

$$s = [M(1 - V\rho/m)]/N_0 f$$

in which $V/m = \langle v \rangle$, the partial specific volume. Substituting,

$$s = [M(1 - \langle v \rangle \rho)/N_0 f]$$

or

$$M = N_0 fs/(1 - \langle v \rangle \rho)$$

From equation (3.1)

$$f = KT/D$$

where f is the frictional coefficient, K the Boltzmann's constant, T the absolute temperature and D is the diffusion coefficient. Substituting,

$$M = N_0 KTs/D(1 - \langle v \rangle \rho) = RTs/D(1 - \langle v \rangle \rho) \tag{3.5}$$

where R is the gas constant and is equal to $N_0 K$. This equation is called Svedberg's equation and is arrived at by combining the sedimentation and diffusion equations. If M and D are known, either s, the sedimentation coefficient, or f, the coefficient of friction, can be calculated. Molecular weight M can also be determined using equation (3.5) if D and s are known. The sedimentation coefficient for any molecule at 20°C and in pure water is taken as the standard for comparison. The expression for this value in terms of the coefficient s at any temperature and in any liquid is given by

$$s_{20,w} = [s\eta(1 - \langle v \rangle \rho)_{20,w}]/\eta_{20,w}(1 - \langle v \rangle \rho)$$

The sedimentation coefficient is measured in Svedberg units ($1S = 10^{-13}$ sec). For example the sedimentation coefficient of one of the subunits of ribosome is 30S and the other is 50S.

3.6 The Ultracentrifuge

The ultracentrifuge is an instrument used to measure sedimentation coefficients. Figure 3.4 shows an ultracentrifuge with its component parts. The sample is held in a centrifugal cell in a titanium or aluminium rotor, which is rotated by an electric motor at speeds up to 70,000 revolutions per minute (rpm). The rotor assembly, including the sample cell, is kept inside an armored chamber to protect the users against any explosive accident. In order to avoid excessive heating that may be caused by

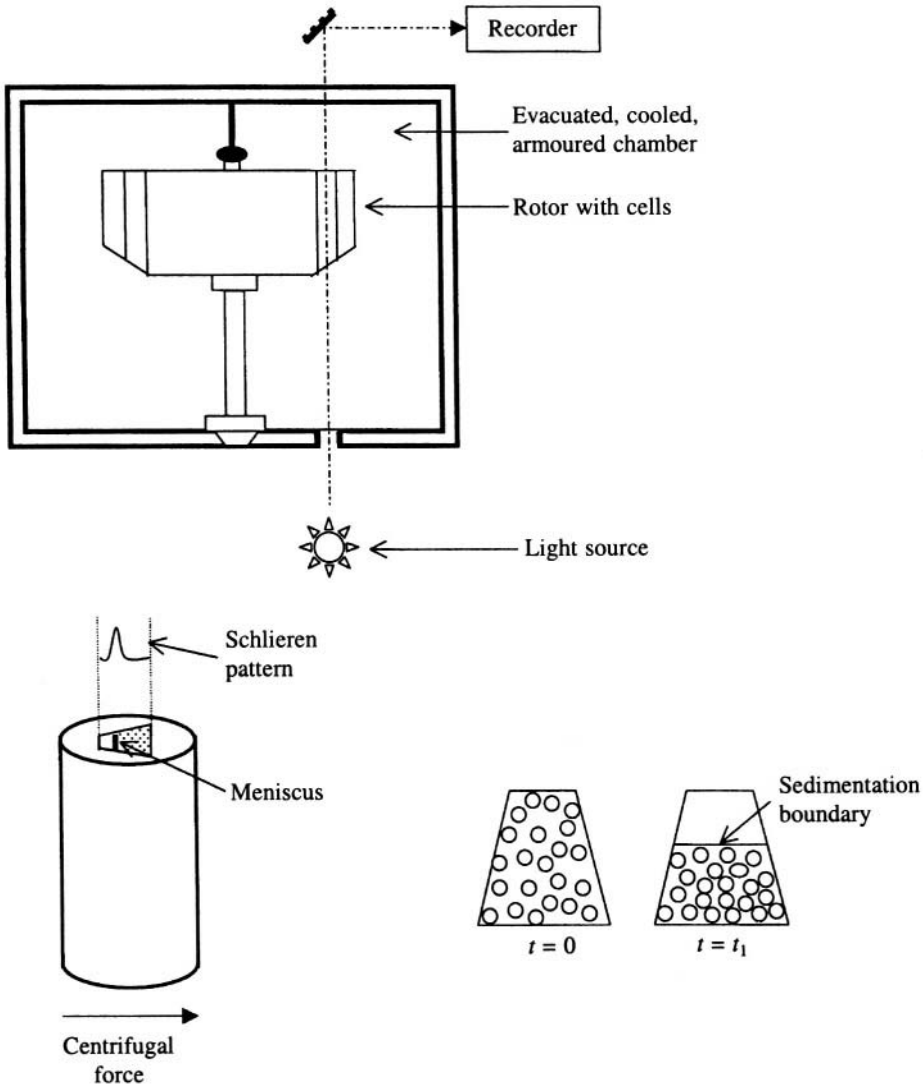


Fig. 3.4 Schematic cutaway sketch of an ultracentrifuge and its parts.

friction with air, the rotor chamber is evacuated and refrigerated. In order to avoid collision of the particles in the sample with the cell wall, the sample cell is shaped like the sectors of a circle drawn around the axis of rotation. Due to the high speeds, the balance of weight on the rotor is critical. A flexible shaft is used which can tolerate a mismatch of up to 0.5 g. A centrifuge can be either a preparative instrument or an analytical one. In a preparative centrifuge the sample is spun for a fixed length of time to separate mixtures or purify samples. In an analytical centrifuge, the movement of the boundary during sedimentation can be monitored using optical devices like Schlieren optics, Rayleigh interference or absorbance, and the period and rate of centrifugation can be accordingly adjusted. The Schlieren optical system is based on the fact that light does not deviate as long as it moves through a solution of uniform concentration but undergoes refraction in a solution with varying densities. The change in refractive index with change in concentration is recorded and is used in the determination of the sedimentation coefficient.

Molecular weight determination can be carried out by the sedimentation velocity technique or the sedimentation equilibrium method. In the sedimentation velocity method the ultracentrifuge is operated at high speeds and the movement of the sedimentation boundary is recorded as a function of time. This is a measure of rate of sedimentation. Rayleigh's interference method is used to make this record, since the Schlieren optical system is not very sensitive to small changes in concentration. The Rayleigh interference system uses a double sector cell, one for the solvent and the other for the solution. By measuring the displacement of interference fringes, the system measures the difference in refractive index between the reference solvent and the solution.

In the equilibrium method the sample is centrifuged till an equilibrium is reached between sedimentation and the diffusive movement of the particles, i.e. there is no net movement of solute particles in the cell and the concentration gradient remains stable. In other words, the net flux J is zero, i.e.

$$J = s\omega^2 rC - D(\delta C/\delta r) = 0$$

or

$$s\omega^2 rC = D(\delta C/\delta r)$$

$$s/D = (\delta C/\delta r)/\omega^2 rC$$

Substituting in the Svedberg equation, we obtain the expression for the molecular weight

$$M = [RT/(1 - \langle v \rangle \rho)\omega^2 rC](\delta C/\delta r)$$

Thus the molecular weight can be determined from a measurement of the concentration gradient. Expressing the above equation in terms of concentration at two different distances from the rotor axis, r_a and r_b we get

$$M = 2RT(C_b - C_a)/[(1 - \langle v \rangle \rho)\omega^2(r_b^2 - r_a^2)C_0]$$

where C_a and C_b are concentrations corresponding to distances r_a and r_b, and C_0 is the initial concentration. The advantage in using this equation instead of the previous one is that r_a and r_b can be measured at the meniscus and at the bottom of the cell respectively. In addition the equation is independent of the size and shape of the molecule. Moreover at the meniscus and at the bottom the equilibrium condition is surely satisfied, i.e. the flux is zero. This equation is, however, valid only for homogeneous samples and therefore the molecule to be analysed will have to be pure. The rotor should be spun at low speeds as otherwise there will be no solute molecules at the meniscus and a

pellet will form at the bottom. Molecular weights ranging from a few hundred daltons to several thousand daltons can be determined using the equilibrium method and the molecular weight decides the rotor speed. For example, a molecular weight of 50,000 daltons will require a speed of 10,000 rpm. The main disadvantage of this method is that long spinning times may be required before equilibrium is reached. Nevertheless, this technique is desirable when the diffusion coefficient D is not known. Another method suggested by Y phantis uses a high speed equilibrium. The centrifuge is operated at such a high speed that the concentration of the solute at the meniscus is zero. Now the difference in refractive index between any point in the cell and meniscus is proportional to the concentration of the solute at that point.

In another technique known as density gradient ultracentrifugation, a solution of low molecular weight (e.g. CsCl) is centrifuged in a cell, so that a density gradient is formed at equilibrium. If we add a substance of high molecular weight to this solution then it will be suspended at a position where its buoyant density is equal to the density of the low molecular weight solution. The position where it is suspended is measured and is used to calculate its molecular weight. This method is widely used in biochemical and biophysical research. If we have a multicomponent system with varying densities then the components will be suspended at different positions in the cell. In practice, a thin layer of the macromolecular sample is layered on a larger volume of the solvent and centrifuged. The components of the system will sediment according to their sedimentation coefficients and separate out as bands (Figure 3.5). This is called zonal sedimentation. The advantage of this technique is that very little **($\sim\mu$g)** sample is used.

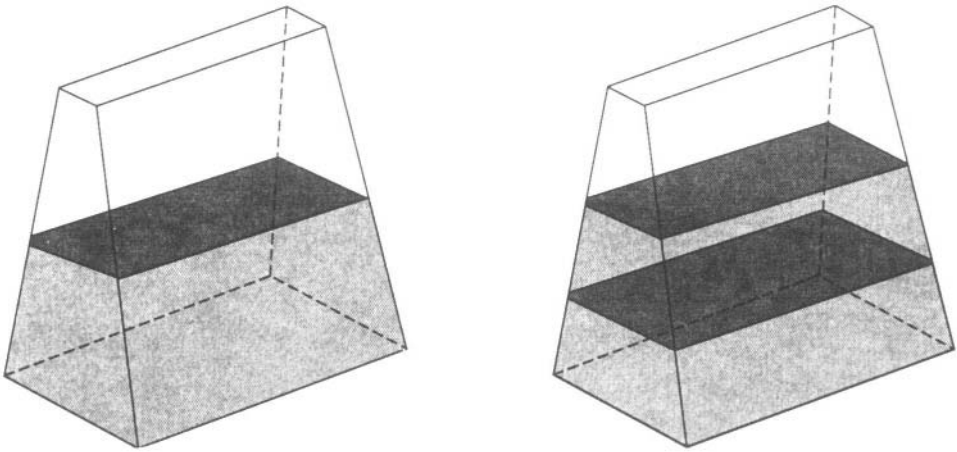


Fig. 3.5 Zone sedimentation

Apart from molecular weights, sedimentation methods can also yield useful information regarding the shape and size of the particles. An equation relating the shape and size to the sedimentation velocity is obtained by manipulating equation (3.4)

$$v_0 = \text{velocity} = [2r_p^2(\rho_p - \rho_m)\omega^2 r]/9\eta$$

where ρ_p and ρ_m are the densities of the particle and the medium respectively. The latter is a correction for the buoyant force. r is the distance of the particle from the axis of rotation, η is the viscosity, ω is the angular velocity. Thus the sedimentation velocity rate is proportional to the radius r_p and to the density of the particle. The equation is true for perfectly spherical particles. In the case of particles

that are not spherical, a correction factor is applied. This is called the shape factor F and is equal to (f_1/f_2), where f_1 is the frictional coefficient of the molecule, and f_2 is the frictional coefficient of an equivalent sphere. Thus r_p, which indicates the size of the particle, and the shape factor, which gives information about the deviation of the particle from sphericity, both may be obtained from sedimentation experiments.

3.7 Viscosity

Viscosity is resistance to liquid flow. It is increased by the presence of a solute. The increase depends on the concentration and other structural properties of the solute and information regarding these properties may be obtained from measurements of viscosity. Viscosity measurement is simple and inexpensive and may be made using a capillary viscometer. A typical example is the Ostwald Viscometer shown in Figure 3.6. The solution whose viscosity is to be determined is poured into the lower bulb through arm A. It is then sucked up through arm B to fill the upper bulb and rise past the mark **a**. Now it is allowed to fall back through the capillary under gravity. The time taken for the solution to fall from point **a** to point **b** is noted. Then η, the viscosity, can be calculated using the following expression, derived from Poiseuille's equation for the viscosity of a liquid flowing through a capillary

$$\eta = (\rho g r^4 t \pi)/(8lV')$$

where ρ is the density of the liquid, g the acceleration due to the earth's gravity, l and r are the length and radius, respectively, of the capillary, t is the time taken by the liquid to fall back from **a** to **b** and V' is a measure of the total hydrostatic pressure to which the liquid flow through the capillary is subjected and is calculated as the integral of dV/h between **a** and **b**, where h measures the height of the liquid column and V is the volume element at the point h. It is difficult to measure the absolute viscosity of a liquid by this method especially since V' is difficult to estimate. However, a comparison with a reference standard can be made. Thus if η_0 is the viscosity of the reference, and ρ_0 its density,

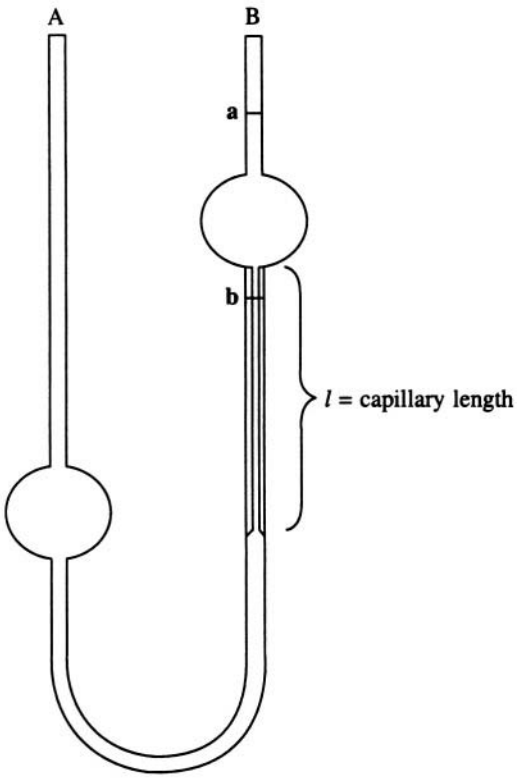


Fig. 3.6 The Ostwald viscometer

$$\eta/\eta_0 = (\rho/\rho_0)(t/t_0)$$

where t_0 is the outflow time for the standard. The reference liquid is usually chosen to be the solvent, which is then used to dissolve the particles of interest in order to measure the viscosity of the solution.

A theoretical treatment of the viscosity of a solution of particles is required before we can obtain shape and size information from viscosity measurements. The viscosity η of a solution, relative to the viscosity η_0 of the pure solvent, can be expressed as a power series

$$\eta/\eta_0 = (1 + k_1C_2 + k_2C_2^2 + \ldots)$$

where C_2 is the solute concentration in g/cm^3. Einstein showed that, ignoring the higher order terms, the above equation may be simplified to obtain the relative viscosity for a dilute solution of large rigid spheres as

$$\eta/\eta_0 = 1+2.5\phi$$

where

$$\phi = (V_hN_0C_2)/M = C_2\langle v \rangle$$

in which V_h is the hydrated volume, (M/N_0) is the mass of the molecule in grams, C_2 the solute concentration in g/cm^3 and $\langle v \rangle$ is the partial specific volume. Thus the change in viscosity on addition of the solute particles does not depend on their size but depends on the volume they occupy in solution.

For particles which are shaped as rigid ellipsoids, an intrinsic viscosity $[\eta]$ may be obtained from the expression

$$[\eta] = (\nu V_hN_0/M)$$

ν is called the Simha factor and contains information on the shape of the molecule. It varies as the ratio of the major and minor axes of the ellipsoids. The above equation can be more generally written as

$$[\eta] = KM^{\alpha}$$

where K and α are constants. Measurements of the intrinsic viscosity together with sedimentation or diffusion measurements can lead to a good estimate of molecular weight by eliminating shape effects. Table 3.2 lists the Simha factors for prolate and oblate ellipsoids with different axial ratios. Thus, for example, rod shaped particles always have a Simha factor greater than 15. Table 3.3 provides some

Table 3.2. Simha and Scheraga-Mandelkern factors for ellipsoids

Axial ratio	Prolate ellipsoid		Oblate ellipsoid	
	ν	$\beta \times 10^{-6}$	ν	$\beta \times 10^{-6}$
1	2.5	2.12	2.5	2.12
10	13.63	2.41	8.04	2.14
30	74.51	2.78	21.58	2.15
100	593.70	3.22	69.10	2.0

Table 3.3. Experimentally determined intrinsic viscosities and molecular weights

Sample	$[\eta]$	$M(cm^3/g)$
Near spherical particles		
Ribonuclease	3.3	13,700
Hemoglobin	3.6	68,000
Rod like particles		
Myosin	217	4,93,000
DNA	5,000	60,00,000

experimentally determined molecular weights using the Simha shape factor. Another factor, which is related to shape of the particles is given by the Scheraga-Mandelkern equation

$$\beta' = \{N_0^{1/3} v^{1/3}\}/\{(162\pi^2)^{1/3} F\} = \{N_{0S}[\eta]^{1/3}\eta\}/\{M^{2/3}(1 - \langle v\rangle \rho)\}$$

where $F = f/f_{sph}$ is the shape factor, f is the frictional coefficient of the particle, f_{sph} the frictional coefficient of an equivalent sphere, N_0 is Avogadro's number, s is the sedimentation coefficient of the solute particles, $[\eta]$ and η are the intrinsic and relative viscosity of the solution respectively, M is the molecular weight of the solute particles, $\langle v \rangle$ is the partial specific volume of the solute and ρ is the density. The values of the Sheraga-Mandelkern factor are tabulated in Table 3.2 as $\beta = 100^{1/3}\beta'$. β is less sensitive to the axial ratio as compared both to the Simha factor v, as well as to another factor called the Perrin factor. For most globular proteins the axial ratio is less than 10 and the β value is close to 2.20×10^6. Using these values, and the Sheraga-Mandelkern equation, molecular weights can be determined to an accuracy better than 10%. Table 3.4 shows some experimentally determined molecular weights using the Scheraga-Mandelkern equation.

Table 3.4. Experimentally determined molecular weights compared with theoretical values from chemical structure

Molecule	From Chemical Structure	From Sheraga-Mandelkern equation	From sedimentation equilibrium
Ribonuclease A (bovine)	13,683	15,200	13,700
Lysozyme	14,211	12,400	14,500
Serum albumin (bovine)	66,296	59,000	68,000

3.8 Rotational Diffusion

Molecules in solution undergo both translational and rotational displacements due to the Brownian motion of the solvent. Each solute molecule undergoes rotation about its centre of mass. This movement is characterized by a coefficient called the rotational diffusion coefficient $[\theta]$ which represents the average angle of rotation per unit time and is dependent on the rotational friction experienced by the particle. For non-spherical particles, a single frictional coefficient is insufficient to describe the rotational friction. In such cases, the frictional forces acting on the particle are described by a tensor[4].

Therefore $[\theta]$ is also a tensor and is given by the equation

$$[\theta] = kT/f_{rot}$$

where f_{rot}, the coefficient of rotational friction, is a tensor. The determination of $[\theta]$ is especially useful in the case of molecules with high axial ratios (length/width) since it gives a measure of the length of the molecule in solution. $[\theta]$ may be measured by a variety of techniques like flow birefringence, electric birefringence and the polarisation of fluorescence. In order to observe rotational diffusion, the equilibrium orientations of the molecule in normal solution must first be perturbed, and then some property of the molecule, which is dependent on the orientation of the molecule, must be

[4] A tensor is matrix of numbers that describes a given property in different directions. In the present instance, the frictional tensor will consist of a 3 × 3 matrix with the diagonal elements giving the frictional coefficients along the three principal axes of rotation, and off-diagonal elements giving the coefficients for rotation about the other axes.

measured.

3.8.1 Flow birefringence measurements

Flow and electric birefringence methods are based on the differences between the refractive index measured in different directions that are produced by oriented molecules in solution. The solution to be studied is placed between two concentric cylinders (Figure 3.7), of which one (generally the outer one) is rotated with respect to the other. This establishes a velocity gradient, which produces a torque on the molecule, and the molecules align their long axis along the flow lines, perpendicular to the radius (Figure 3.8). The higher the axial ratio of the molecules, the better the alignment. Now a parallel beam of light is passed through a polariser[5], and then through the solution, and then through an analyser. To start with, as the molecules are randomly oriented, the solution is optically isotropic and all the incident light is blocked by the analyser. When the outer cylinder is rotated, the molecules become oriented as described above and the solution becomes birefringent, i.e. the refractive index of the solution is not the same in all directions. The light transmitted through the solute now becomes elliptically polarised and hence some of it will pass through the analyser. An image of the analyser appears bright everywhere excepting in four perpendicular regions where the optic axis of the molecule is parallel to the plane of polarisation of the incident light (Figure 3.7). These four regions form a dark cross, corresponding to regions of zero birefringence. The cross is seen initially at 45° to the polariser

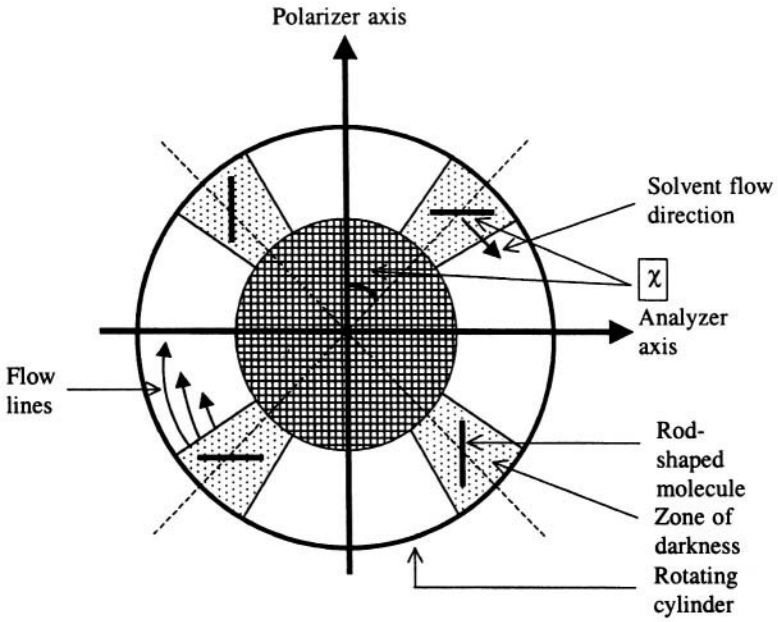
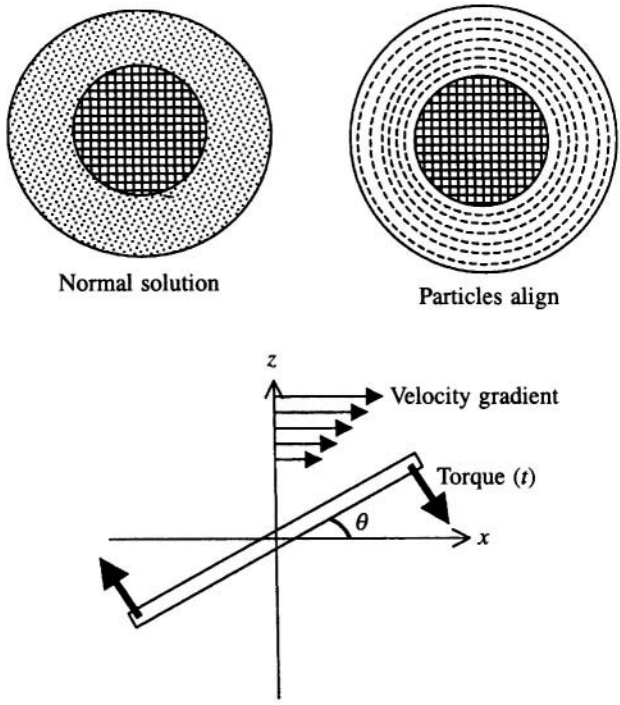


Fig. 3.7 Apparatus for the measurement of flow birefringence.

axis when there is a only slight degree of orientation of the molecules. As the degree of orientation increases, the dark cross rotates to align itself with the cross formed by the polariser and analyser.

The position of the cross is measured by the extinction angle χ(Figure 3.7). A plot of χ versus

[5]See Chapter 5 for an elementary discussion of the polarisation of light.



At $\theta = 0°$, t is minimum; at $\theta = 90°$, t is maximum, where t is the torque causing the molecules to rotate in the clockwise direction.

Fig. 3.8 Alignment of particles in flow birefrigence

shear gradient G is made to determine $[\theta]$. The shear gradient depends on the speed of the outer cylinder, and the distance between the cylinders. Therefore χ, which is the angle between the macroscopic optic axis of the solution of oriented molecules and the direction of the solvent flow, is a measure of the degree of orientation. Rod-shaped molecules tend to orient as shown in Figure 3.7, with their long axis parallel or perpendicular to the polariser axis. The measurement of $[\theta]$ then enables the calculation of the frictional tensor using the equation above.

3.8.2 Electric birefringence

Instead of a velocity gradient, an electric field can be used to orient the sample. Electric birefringence is then an optical measurement of the orientation caused by the electric field. An electric pulse, generally a square wave, is applied to the solution for a sufficiently short duration such that there is no electrophoretic mobility, only rotational reorientation of the molecules (Figure 3.9). The macromolecules in the solution get preferentially aligned in a particular direction in the electric field. The degree of alignment depends on the electric properties of the molecule and is measured by the difference between the refractive index of the solution parallel to the applied field and the one perpendicular to it. This is known as form birefringence. Intrinsic birefringence refers to the difference in the refractive index between two perpendicular directions within a rod-like molecule such as DNA. The experiment to measure the electric birefringence consists of passing light through crossed polarisers, with the sample in the electric field placed in between. The light is finally received by a photomultiplier

tube connected to an oscilloscope. Since the molecules are oriented by the electric field, a static measurement of the birefringence at the equilibrium state will yield little information about the particle's shape. However, the decay in the birefringence, when the electric field is removed will depend on the molecular shape. This is measured as the decay of the birefringence signal and the rotational diffusion coefficient of the molecule is obtained.

3.9 Light Scattering

Light scattering is a technique widely used to determine the molecular weight of macromolecules in solution. A beam of light of a particular wavelength illuminates the sample. The experiment is sensitive to impurities hence great care must be taken to exclude dust particles from the solution. Radiation scattered by the sample, at the same wavelength is measured as a function of the angle between the incident beam and the detector. The chief difference between light scattering and visible range spectroscopy is that in the former technique, only elastic scattering by the particles in the solution is considered. This is the reason the measurement of scattering is made at the same wavelength as the incident beam. In this respect, the technique is similar to diffraction. However, in diffraction, we measure the interference patterns formed by the radiation scattered from different particles. In light scattering the solution is dilute enough to consider that the molecules scatter independently and are randomly oriented. Moreover they do not absorb the incident radiation, and therefore the scattered radiation has the same wavelength. Hence the term elastic scattering. In general, there are two

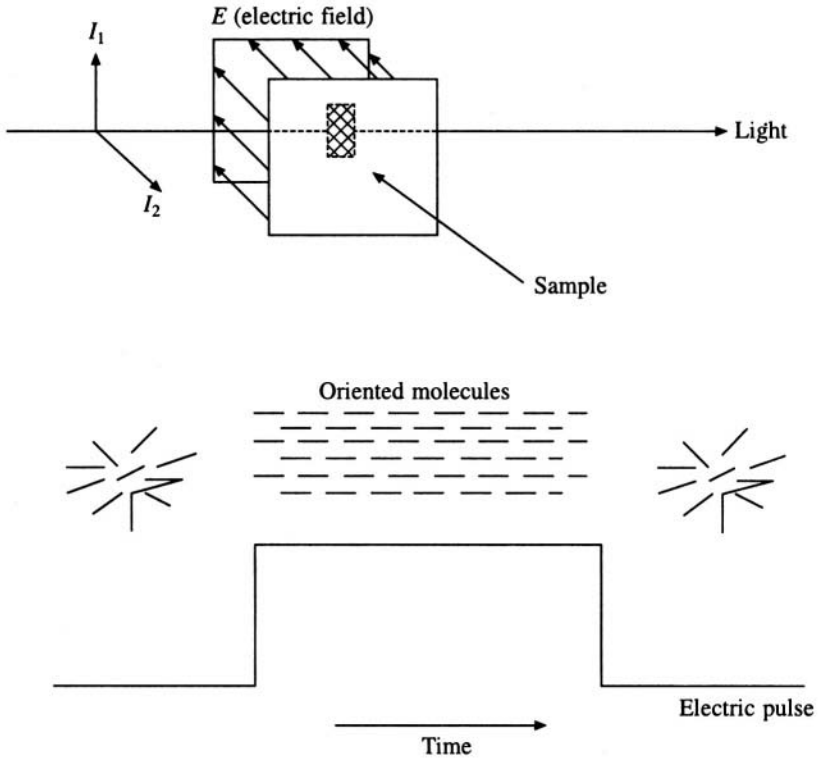


Fig. 3.9 Electric dichroism. The sample is placed in an electric field. On application of an electric pulse, the molecules in the sample become oriented along one direction. Removal of the electric field results in the molecules reverting to a disoriented state

situations: (a) when the particles are smaller than the wavelength, and (b) when they are comparable in size to wavelength.

Consider first the situation when the molecules are smaller than the wavelength of the incident and scattered radiation. Since light is electromagnetic radiation, it has electrical as well as magnetic components. But for the purposes of the present discussion we concentrate only on the electric field component *E,* and ignore the magnetic component. If the radiation is of frequency *v,* and the maximum amplitude is E_0 then

$$E(t) = E_0 \cos (2\pi\nu t)$$

where *t* is the time. According to classical electromagnetic theory, when light strikes matter, electrons are displaced from their equilibrium positions with respect to the nucleus and a dipole is formed. The dipole moment *p* so induced is proportional to the electric field and is given by

$$p = \alpha E = \alpha E_0 \cos (2\pi\nu t)$$

where α is a constant of proportionality called polarisability. The polarisability is a measure of the ease with which an electron can be displaced from its equilibrium position. Since the incident electric field is oscillating constantly, the dipole also oscillates in magnitude, going from zero to a maximum and back. Again from electromagnetic theory, an oscillating dipole gives rise to scattered radiation, whose electric field is given by

$$E_s = [4\pi^2\nu^2\alpha^2 E_0 \cos (2\pi\nu t)]/c^2 r$$

in which *c* is the velocity of light and *r* the distance of the dipole from the detector. The ratio of the scattered radiation I_s to the incident intensity I_0 is

$$I_s / I_0 = E_s^2 / E_0^2 = (16\pi^4\alpha^2)/\lambda^4 r^2$$

in which we have substituted $\nu = c/\lambda$. Thus the ratio is inversely proportional to the fourth power of the wavelength λ. This is called Rayleigh scattering. If, instead of unpolarised or circularly polarised light, we use light polarised along a particular axis then the induced dipole oscillates only in that direction. As a consequence the scattered radiation will also be polarised and the above equation has to be modified for such effects. Thus

$$I_s/I_0 = [(8\pi^4\alpha^2)/\lambda^4 r^2](1 + \cos^2\theta)$$

for a single scattering particle. If there are *N* particles, then

$$I_s/I_0 = N[(8\pi^4\alpha^2)/\lambda^4 r^2](1 + \cos^2\theta)$$

The Rayleigh ratio R_θ is the relative intensity of the scattered light and is given by

$$R_\theta = [r^2/(1 + \cos^2\theta)V][I_s/I_0]$$
$$= (8\pi^4\alpha^2 N')/\lambda^4$$

where $N' = N/V$ is called the number density of particles and is equal to CN_0/M (where *C* is the concentration, N_0 is Avogadro's number and *M* the molecular weight), and *V* is the volume of the solution. When we substitute for the polarisability α in terms of refractive index and concentration of the solution, we obtain

$$R_\theta = KMC$$

where K is a constant for a given solution incorporating the refractive index of the pure solvent, Avagadro's number, the incident wavelength and the rate of change of refractive index with concentration. Thus if R_θ is measured, and the concentration C is known, the molecular weight can be determined. When the sample is a heterogeneous mixture of several molecular species of concentrations C_1, C_2... C_i, then, instead of a single molecular weight, the light scattering experiment will yield

$$\Sigma_i M_i C_i$$

where M_i is the molecular weight of the component i. The weighted average molecular weight is then obtained as

$$\langle M_w \rangle = (\Sigma_i M_i C_i)/\Sigma_i C_i$$

The other situation is when the molecules in the solution have a size, which is comparable to the wavelength of the incident radiation. In the previous case, when the particle size was small, we could assume that the dipoles were oscillating in phase for all the scattering elements within the particle. However, in the present case, different parts of the molecule can be considered to scatter independently and thus the radiation emanating from the different portions undergo interference (Figure 3.10). This effect is taken care of by applying a correction factor P which is a function of θ to the Rayleigh scattering ratio. $P(\theta)$ can be expressed in terms of radius of gyration R_G of the molecule. The radius

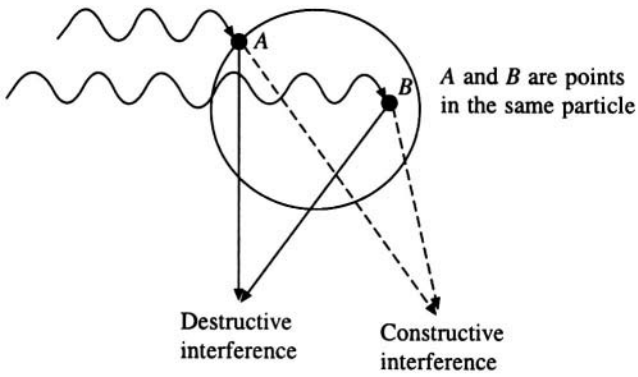


Fig. 3.10 Interference in light scattered from particles with size comparable to λ, the wavelength of the incident light

of gyration gives information about the shape of the molecule. Thus

$$R_\theta = KMCP(\theta) = KMC(1 - 16\pi^2 R_G^2 \sin^2\theta/3\lambda^2)$$

Or at the limit of infinite dilution (and since $1/(1 - x)$ is approximately equal to $(1 + x)$ for small values of x),

$$(KC/R_\theta) = (1/M)(1 + 16\pi^2 R_G^2 \sin^2\theta/3\lambda^2)$$

KC/R_θ is measured for various values of the scattering angle θ, and plotted against $\sin^2(\theta/2) + k'C$, where k' is an arbitrary constant used for scaling (Figure 3.11). This is known as the Zimm plot. The concentration data at each angle is extrapolated to $C = 0$. Similarly the angle-dependent scattering data is extrapolated to $\theta = 0$. Both the extrapolations intersect on the y axis giving $1/M$. The radius

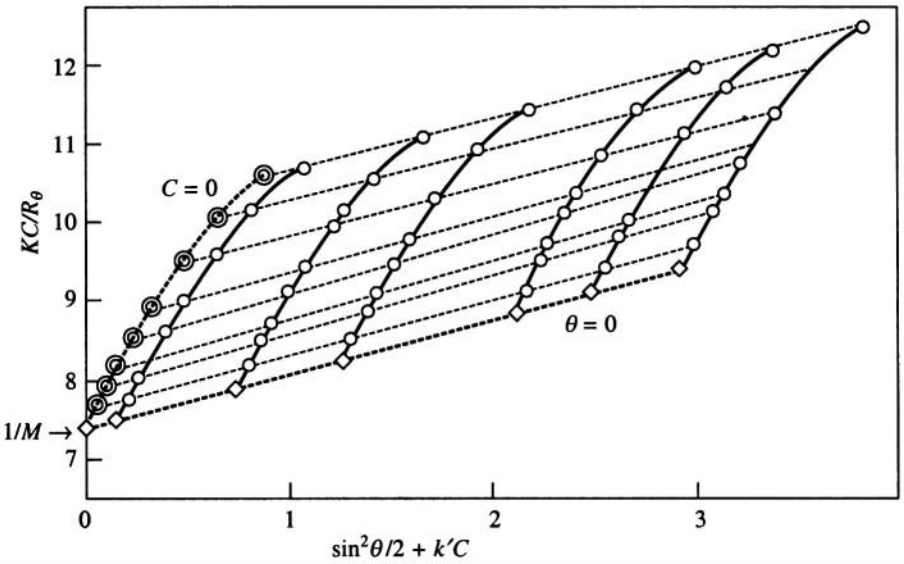


Fig. 3.11 Zimm plot for light scattering from a biomolecular solution. Experimental data (○), extrapolations to zero concentration (⊙) and extrapolations to θ = zero (◇), k′ is a constant

of gyration can be obtained from the slope of the curve C = 0. The value of R_G depends on the shape of the particle. For a sphere of radius *r,* the radius of gyration is given by $R_G^2 = (3/5)r^2$, whereas for an ellipsoid of semi-axes *a* and *b,* $R_G^2 = (a^2 + 2b^2)/5$. For a long rod of length *l,* $R_G^2 = l^2/12$. In this manner one can obtain information about the molecular weight and the radius of gyration, and therefore the shape of the molecule, from the light scattering experiment.

3.10 Small Angle X-ray Scattering

This technique is analogous to that of light scattering. For an average small protein having a maximum dimension of 100 Å, light scattering can give information about the molecular weight but not about the shape of the protein, since the correction term $P(\theta)$ is close to unity. In order to obtain information on the shape of these molecules, radiation with a smaller wavelength, like X-rays, need to be used. Small angle X-ray scattering or SAXS is a diffraction technique. In order to understand the need for measurements at small angles, from the diffraction condition or Bragg's law, the angle of diffraction for an inter-planar spacing of 100 Åcan be calculated to be

$$\sin\theta = \lambda/(2 \times 100) = 0.0077$$

for CuK_α radiation of wavelength 1.5418 Å and $\theta \sim 0.45°$. For a spacing of 1000 Å, θ is 0.045°. Thus, the angle (with respect to the direct beam) at which the measurements will have to be made to obtain information of this nature are very small. The experiment consists of allowing a narrow X-ray beam of known wavelength to be incident on a dilute solution of the sample, and then measuring the scattered intensity as a function of the angle. Since the scattering angle of interest is only a few degrees, at the most, from the direct beam, great care must to taken to collimate the beam to make it as thin as possible. Also in the measurement of the scattered intensity, the direct beam must be carefully and completely excluded. As in light scattering experiment, since the solution is sufficiently dilute, we assume that the molecules scatter independently. For each molecule, the molecular structure

factor is defined as the sum of individual scattered X-rays from all the atoms in the molecule. Taking into consideration the difference in scattering between the different relative orientations of neighbouring molecules and integrating the structure factor over all possible relative orientations, we obtain the expression for the scattering intensity as

$$\langle I(\theta) \rangle = n_\varepsilon^2 (1 - 16\pi^2 R_G^2 \sin^2 \theta / 3\lambda^2)$$

for small values of θ, where R_G is the radius of gyration and n_ε is the number of electrons in the molecule (as obtained from a knowledge of the molecular weight). This equation indicates that the zero-angle X-ray scattering from the solution is approximately proportional to the square of the molecular weight of the particle. A plot of the intensity versus $\sin^2 \theta$ yields the radius of gyration. A more common way of using this expression is to rewrite it as

$$\ln [I(\theta)/I(0)] = -(4\pi R_G \sin \theta/\lambda)^2/3$$

Now the small angle scattering can be plotted as $\ln [I(\theta)/I(0)]$ versus $\sin^2 \theta/\lambda^2$. This plot is called the Guinier plot and it yields the radius of gyration even in the absence of information about the molecular weight. R_G in turn gives information on the shape of the molecule.

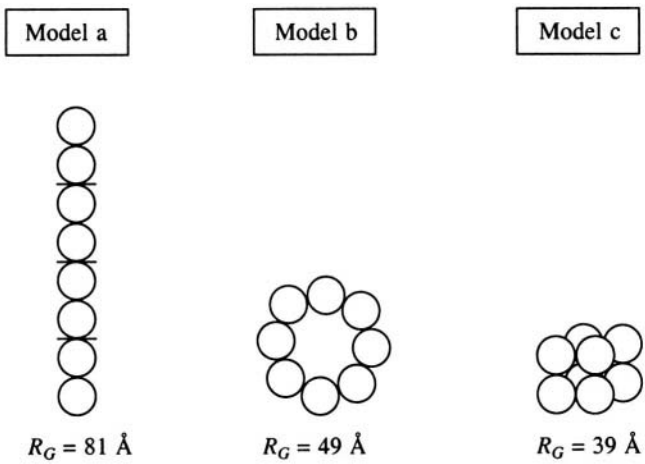


Fig. 3.12 Models for the quarternary structure of α-lactalbumin at pH 4.5

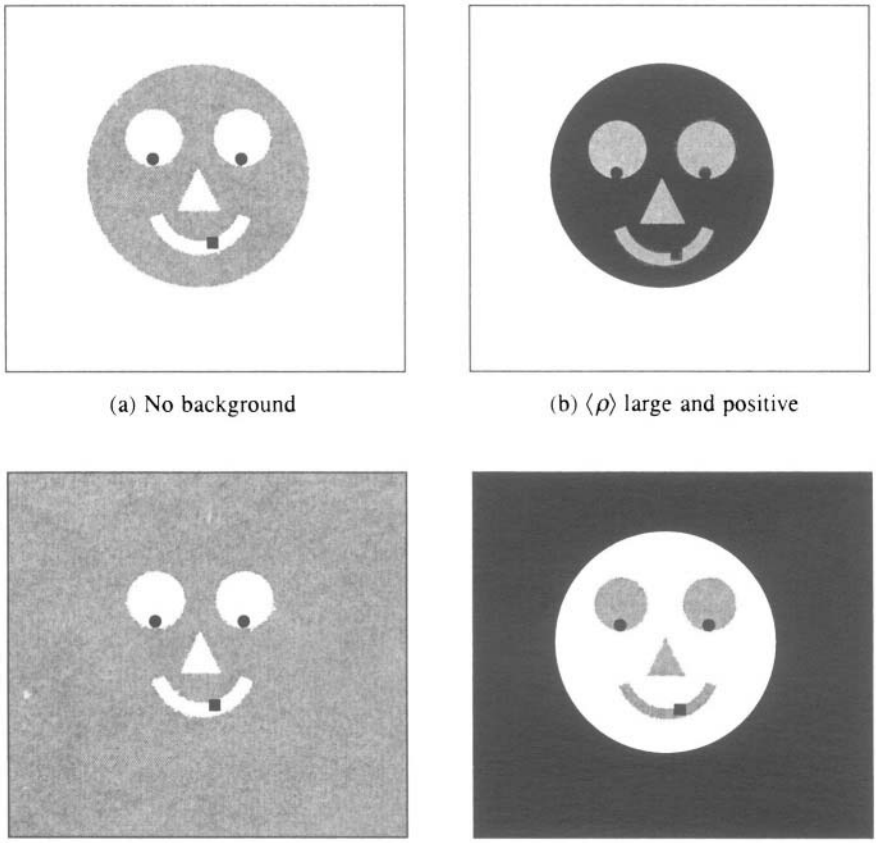
As an example of the application of SAXS consider the experiments on **α-lactalbumin.** At pH 4.5 these gave the radius of gyration as 34 Å. From previous experiments it was known that this protein exists in solution as a tetramer of four dimeric units. Various models for the arrangement of these units (Figure 3.12) were considered and the radius of gyration calculated for each was compared with the experimental value. It is clear that model c in Figure 3.12 is likely to be the correct one.

The scattering of neutrons from solutions can also be used to analyse the shape and size of molecules in solution. The techniques and the theory are similar to SAXS. The major difference is that neutrons are scattered by the atomic nucleus, while X-rays are scattered by the electrons around the nucleus. Therefore, the two techniques may be used in a complementary fashion. In SAXS and other diffraction experiments using X-rays, the intensity of the radiation scattered is a measure of the electron density in the sample and is proportional to the atomic scattering factor. On the other hand, the intensity of neutron scattering is proportional to a factor called the scattering length of the atomic

nucleus. Each element has a different value for the scattering length (Table 3.5). This value depends on the precise nature of the interactions of the incident neutrons with the nucleus. It therefore does not show any relationship with the atomic number or atomic weight. This is unlike the case for X-ray scattering, which is related to the atomic number. The scattering length is a measure of the scattering

Table 3.5. Values of the neutron scattering lengths for some biologically relevant elements

Element	Neutron scattering length ($b \times 10^{-13}$ cm)
H	-3.74
D	6.67
C	6.65
N	9.40
Fe	9.51
Pt	9.50



(a) No background

(b) $\langle \rho \rangle$ large and positive

(c) $\langle \rho \rangle \approx 0$

(d) $\langle \rho \rangle$ large and negative

Fig. 3.13. Efect of contrast in neutron scattering

power. Neutron scattering lengths can be calculated from absolute scattering intensity measurements. Note that the scattering length for hydrogen is negative and of a magnitude comparable to heavier atoms such as carbon. Thus unlike as in the case for X-rays, the contribution of hydrogen atoms to neutron scattering is clearly distinguishable. Neutron scattering can give supplementary information on hydrogen atoms which is not available from X-ray scattering.

Another interesting feature is that the neutron scattering from deuterium is positive and of higher magnitude compared to scattering from hydrogen. By replacing different proportions of water (H_2O) with deuterated water (D_2O) in the solvent, its scattering density can be varied at will. In this way the contrast between solute and solvent can be manipulated to observe details of interest. The constrast $\langle\rho\rangle$ is the mean difference in scattering density between the solute and the solvent (Figure 3.13). Thus,

$$\langle\rho\rangle = \langle\sigma_{\text{macromolecule}}\rangle - \langle\sigma_{\text{solvent}}\rangle$$

where $\langle\sigma\rangle$ represents the average neutron scattering intensity of the appropriate component of the solution. When there is no background, the object will appear as in Figure 3.13(a). When $\langle\rho\rangle$ is positive and large, the object will appear as in Figure 3.13(b). When $\langle\rho\rangle$ is zero then the shape of the object will not be seen (Figure 3.13(c)) but internal details will be visible. When $\langle\rho\rangle$ is large and negative we see the object as in Figure 3.13(d). Thus contrast plays an important role in determining the amount of detail that can be obtained.

Deuterium can also be covalently substituted for hydrogen in cases where the sample is composed of two components. One of the components is labeled with *H* and the other with *D*. Then by adjusting the H_2O–D_2O composition of the solvent, the situation which can be arrived at is

either $$\langle\sigma_1\rangle \equiv \langle\sigma_{\text{macromolecule}}\rangle$$

or $$\langle\sigma_1\rangle \equiv \langle\sigma_{\text{solvent}}\rangle$$

and the scattering due to one of the components will be blacked out. For example, when dealing with icosahedral viruses, the protein coat encapsulating the nucleic acid molecule of viral genome usually makes it difficult to observe the latter in scattering experiments. If we consider that the neutron scattering length per unit volume for fully protonated protein is 3.1×10^{14} cm/Å^3, while that of nucleic acid is 4.4×10^{14} cm/ Å^3, and that for fully deuterated protein it is 8.5×10^{14} cm/Å^3, and for fully deuterated nucleic acid it is 7.4×10^{14} cm/ Å^3, we see that by selective labeling, the nucleic acid core and the protein coat of viruses can be observed independent of each other.

4

Spectroscopy

4.1 Introduction

Electromagnetic energy depends on the wavelength (or frequency) of the radiation. Different wavelengths are useful in probing different aspects of materials. Table 4.1 indicates how the various regions of the

Table 4.1 The electromagnetic spectrum and the corresponding physical experiments

Name	Wavelength (meters)	Frequency (Hz)	Use
X-rays	$10^{-12} - 10^{-8}$	10^{20}–10^{16}	Diffraction, small angle scattering
Ultraviolet (UV)	$10^{-8} - 4 \times 10^{-7}$	$10^{16} - 7.5 \times 10^{14}$	Electronic structure of molecules
Visible	$4 \times 10^{-7} - 7.5 \times 10^{-7}$	$7.5 \times 10^{14} - 4 \times 10^{14}$	Electronic structure of molecules
Infrared (IR)	$7.5 \times 10^{-7} - 10^{-3}$	$4 \times 10^{14} - 10^{11}$	Vibrational and rotational spectra of molecules
Microwave	$1 \times 10^{-3} - 1$	$10^{11} - 10^{8}$	Rotational spectra
Radiowaves	$1 - 10^{3}$	$10^{8} - 10^{5}$	NMR

electromagnetic spectrum are used.

In a spectroscopic experiment electromagnetic radiation of a particular frequency or a range of frequencies is allowed to fall on the sample under study. The radiation coming out of the sample is then analysed in terms of the intensity at different frequencies. This indicates which particular frequencies are absorbed (or emitted) by the molecule, thus giving a picture of the molecular energy levels. These, in turn, can be interpreted in terms of the chemistry or stereochemistry of the molecule. Under ordinary circumstances, all the atoms and molecules would exist in the ground state energy level. When, due to the incident radiation, electrons in the outer shell are raised by one energy level, the phenomenon is known as resonance absorption and the technique of atomic absorption spectroscopy

is based on this. The energy required lies in the ultraviolet region of the electromagnetic spectrum. Molecular spectra have additional complexities arising from vibrational and rotational motion. In this case, transitions between electronic levels are found in the ultraviolet and visible regions, those between different vibrational levels but within the same electronic level are in the infrared region, and those between rotational levels are in the infrared and microwave regions. Ultraviolet absorption spectra in molecules appear as wide bands. This is because the molecule usually contains sets of closely spaced electronic energy levels, and the transitions between these levels, as indicated by the spectra, are difficult to resolve. Fluorescence Spectroscopy of molecules also makes use of the absorption phenomenon but in a different way. A part of energy gained by a molecule by absorption of electromagnetic radiation may be dissipated away, for example, as heat. The remaining portion of the energy is radiated as the molecule returns to its ground state. Since the energy now is not as much as was originally absorbed, the frequency of emitted radiation is less than that of the incident radiation. This process is called fluorescence. For example, many organic compounds fluoresce in the visible region after being irradiated with ultraviolet radiation. Many molecules have the ability to rotate the plane of polarisation of plane-polarised light. This property is called optical activity. Its variation with wavelength of the light is called optical rotatory dispersion. A related phenomenon is circular dichroism, where the polarised light is split into left and right circular polarisation and the difference in the absorbance of the two components is measured. All these spectroscopic techniques give a greater or lesser amount of information about the stereochemistry and conformation of the molecules in the sample. They are of great utility in studies of biological molecules. We will now treat each of them in greater detail.

4.2 Ultraviolet/Visible Spectroscopy

If a beam of light of intensity $\mathbf{I_0}$ falls on a cell containing the sample, the emergent radiation intensity I is less intense than the incident radiation, since a portion of it is absorbed by the sample. The absorption is different at different wavelengths and is characteristic of the sample. This characteristic quantity is called the absorbance and may be calculated from the Beer-Lambert principle. The Beer-Lambert law states that in a sample, each successive portion along the path of the incident radiation, containing an equal number of absorbing molecules absorbs an equal fraction of the radiation that traverses it. If we consider an infinitely thin slice of the sample *(dl),* the light traversing it may be considered a constant. Then the fraction of the light absorbed is simply proportional to the number of absorbing molecules,

$$-dI/I = C\varepsilon' \, dl$$

where *C* is the concentration of the molecule in moles per litre, ε' is a proportionality constant called the molar extinction coefficient and the negative sign indicates a diminution of the intensity of the radiation as it traverses the sample.ε' is a constant for a given wavelength for a given molecule and is independent of the concentration. Integrating this equation over the entire sample thickness 1, we obtain

$$A = \log (I_0/I) = C\varepsilon(\lambda)l \qquad (4.1)$$

where $\varepsilon = \varepsilon'/2.303$, i.e. molar extinction coefficient converted to log base 10. It is expressed as a function of wavelength λ. The term $\log(I_0/I))$ is called the absorbance A. In an experiment, if *A* is measured at each wavelength λ, $\varepsilon(\lambda)$ can be calculated and tabulated. Usually, the wavelength λ_{max} at the maximal extinction and the extinction coefficient ε_{max} at this wavelength are used to characterise the sample. Sometimes, especially for biopolymers, the molar extinction coefficient is difficult to

measure. Instead the average extinction per residue is calculated. In the actual experimental set up, in order to eliminate the effect of the solvent, the effect of the path length, etc., what is measured usually is the difference in absorbance between a cell containing the sample, and one containing the solvent alone. A double beam spectrophotometer has a beam splitter that splits the beam into two equidistant paths. The sample is placed in the path of one half of the beam and the reference in the other.

UV/Vis absorption spectroscopy is used frequently in biophysical research for simple tasks such as determining the concentration of a sample, as also for answering complex structural questions. Electronic transitions between molecular orbitals are responsible for the spectrum and each spectrum characterises the quantum mechanical state of the molecule. The most commonly observed transitions in biomolecules are Between the: (a) ground state π orbital and the excited π^* orbital, i.e. $\pi \rightarrow \pi^*$ transition and (b) nonbonding n orbital and the π^* orbital, i.e. $n \rightarrow \pi^*$. Though the σ orbitals do not usually participate, $n \rightarrow \sigma^*$ transitions have also been observed. At any given wavelength, one particular chemical group in the molecule will dominate the absorption spectrum. Such groups are called chromophores. In biological molecules such as nucleic acids and proteins the chromophores absorb light at wavelengths less than about **300 nm**[6]. Biologically interesting macromolecules cannot be studied in the gas phase and the presence of the solvent itself becomes a constraint since the chromophores of the solvent will dominate at certain wavelengths. Thus, the most common solvent, water, absorbs very strongly below 170 nm, restricting useful study to wavelengths above this value. Protein molecules have three types of chromophores, namely the peptide group, the amino acid side chain groups and any prosthetic groups that may be present. The peptide group has an $n \rightarrow \pi^*$ absorption band at 210–220 nm. The intensity of this band is much smaller than the intense $\pi \rightarrow \pi^*$ transition at 190 nm, which may be considered a 'signature' absorption of the peptide group. A third transition corresponding, tentatively, to the $n \rightarrow \sigma^*$ transition in the peptide, is also sometimes observed. As far as the amino acid side chains are concerned, many of them absorb in the same region as the peptide group. In the case of the aromatic amino acids tryptophan, tyrosine and phenylalanine, the dominant band is between 230 and 300 nm, i.e. above the peptide bands. The most intense is the 280 nm absorption of tryptophan. Tyrosine also has an absorption band at almost the same wavelength, though somewhat weaker. Phenylalanine is much weaker and at approximately 260 nm. It is usually observed only in the absence of significant amounts of tryptophan or tyrosine in the protein. The absorption bands of the side chains change with the pH of the solvent. This is mainly due to the change in the protonation of the chromophores. For example, the removal of the OH proton in tyrosine at pH > 10.9 shifts the absorbance band from 274 nm to about 295 nm. Disulphide formation changes the spectrum of cysteine, but the absorption is weak and difficult to observe. The other possible chromophores in protein arise from the prosthetic groups such as the heme group, flavin, etc. Routine measurements of the concentration of protein solution assume that 1 mg per ml of protein will have an absorbance of 1.0 at 280 nm, i.e. $A_{280} = 1$. This number was arrived at after a series of experiments on proteins with a known number of tyrosines and tryptophans. If the measurement is made at 230 nm, $A_{230} = 3$ for a solution containing 1 mg of protein per ml of solution.

The UV absorption of nucleic acids is almost entirely due to the bases. The sugar-phosphate backbone has little or no contribution at wavelengths greater than 200 nm. The spectra of the individual bases are quite complex, but are all marked by a strong, broad band near 260 nm. Each of the bands corresponds to several $\pi \rightarrow \pi^*$ and $n \rightarrow \pi^*$ transitions. Unusual bases such as the *Y* base ($\lambda_{max} = 325$ nm) and 4-thiouridine ($\lambda_{max} = 340$ nm) found in tRNA have the absorption maxima far away from the

[6]nanometre = 10^{-9} metre

positions for the normal bases. This leads to their use in monitoring tRNA structure and interactions. Nucleic acid polymers show dramatic hypochromism and hyperchromism. These are effects that arise due to aggregation of chromophores. Quantum mechanical theory says that due to interactions between the electric dipoles present in the chromophores, groups of chromophores, which aggregate together, have different effects on the absorption spectrum. In a polynucleotide, if the bases are stacked one on the other, as for example in the double helical structure, this leads to a diminution of the intensities of absorption. The effect is most noticeable at the 260 nm band and is called hypochromism. If on the other hand the bases are arranged in almost the same plane, i.e. in an end-to-end arrangement, the absorption intensity is increased leading to hyperchromism (Figure 4.1). Hyperchromism is commonly used to follow the denaturation behaviour or the melting behaviour of the DNA double helix. At higher temperatures or at other disruptive conditions, when the base-base stacking is disturbed and the double helix is denatured, the intensity of absorption at 260 nm increases.

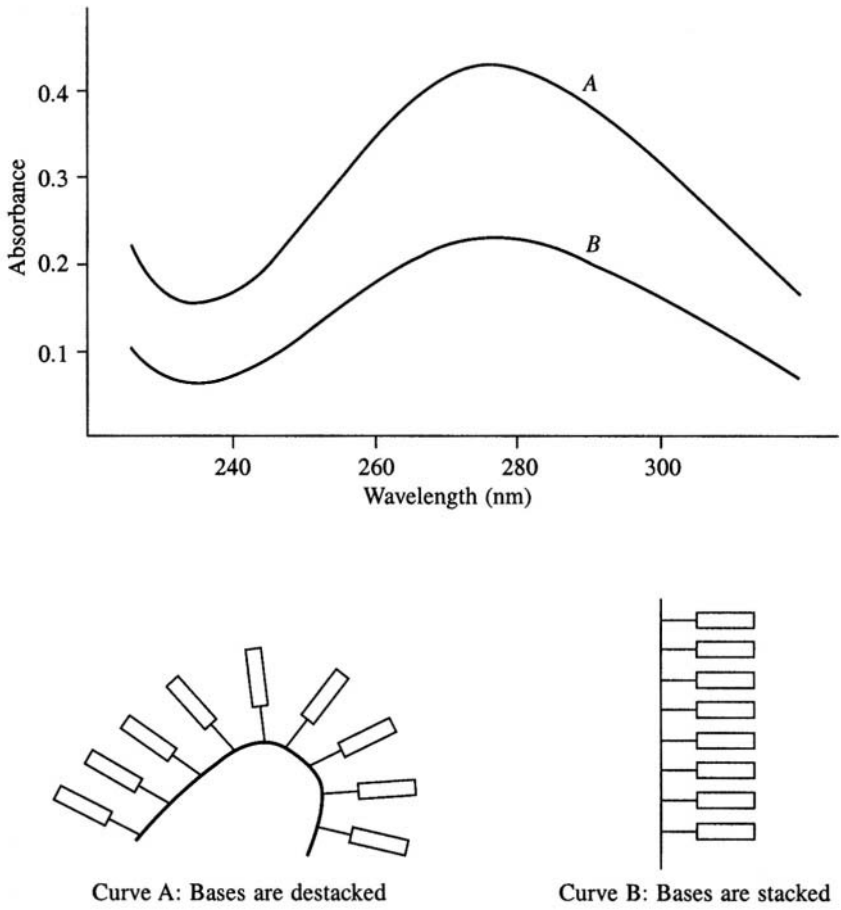


Fig. 4.1 The hyperchromic A and B effects in nucleic acids.

4.3 Circular Dichroism (CD) and Optical Rotatory Dispersion (ORD)

Most biological molecules or systems are optically active; i.e. they rotate the plane of polarised light. If the rotation is clockwise when seen looking towards the light source, it is said to be dextro-rotatory.

If the rotation is counter clockwise, it is levo-rotatory. For any compound, the extent of rotation depends on the number of molecules in the path of the polarised light, or in other words, for solutions, on the concentration and on the path length of the beam through it. It also depends on the wavelength of the radiation and the temperature. Optical activity is quantified by the specific rotation

$$[\alpha]_{\lambda}^{t} = \frac{\alpha^0}{dc}$$

i.e. the specific rotation $[\alpha]$ at a temperature t and for wavelength λ is the measured rotation α^0 divided by the concentration c in grams per cubic centimetre and the light path d in centimetres. Optical activity may also be quantified as the molar rotation

$$[M]_{\lambda}^{t} = \frac{\alpha M}{dc}$$

where M is the gram molecular weight of the substance. For polymers, it is common to use the mean residual rotation

$$[m]_{\lambda}^{t} = \frac{\alpha M_0}{dc}$$

where M_0 is the mean residue molecular weight, i.e. the average molecular weight of the monomers. Rather than the rotation at a single wavelength, measurement of the change in optical activity with change in wavelength yields more useful structural information. This is called Optical Rotatory Dispersion (ORD). A closely related phenomenon is Circular Dichroism (CD). Plane polarised light can be resolved into two beams, each circularly polarised, but in opposite senses (Figure 4.2). The

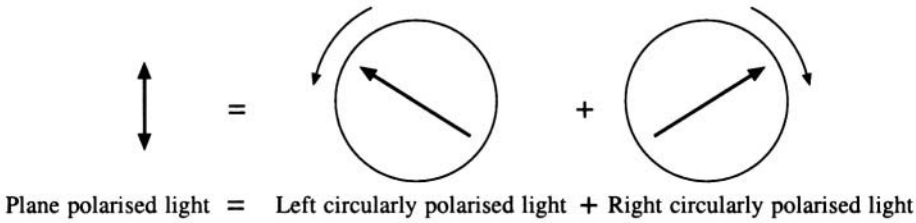


Fig. 4.2 Plane polarised light can be resolved into two components

refractive index for the right circularly polarised light, n_R may be different from that for the left circularly polarised light n_L. Similarly the absorbance A_R for the right circularly polarised light may be different from A_L. When n_R and n_L are unequal, each component will be retarded to a different extent as it passes through the sample. This leads to a change in the phases of the two components when they make up the plane-polarised light, leading to a rotation of the plane of polarisation. This is in fact the theoretical basis for the phenomenon of optical activity. Since the refractive index is always dependent on the wavelength, the angle of rotation is also wavelength-dependent. At a given wavelength λ, the angle of rotation

$$\alpha_{\lambda} = (180d/\lambda)(n_L - n_R)$$

where d is the path length. α, or more commonly, the specific rotation $[\alpha]_{\lambda}$ can be plotted as a function of wavelength (Figure 4.3). This is called the Optical Rotatory Dispersion (ORD) spectrum. The difference in the absorption between the right and the left components does not introduce a phase

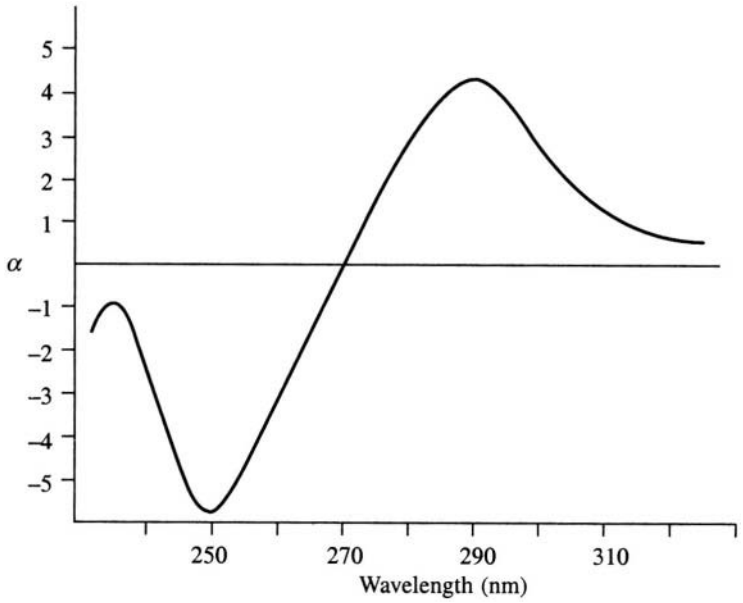


Fig. 4.3 A typical ORD spectrum.

difference between the two, but instead produces a difference in intensity, i.e. a difference in magnitude of the wave. The resultant beam is no longer plane polarised but is elliptically polarised (Figure 4.4). The difference in absorbance $A_L - A_R$ is called the circular dichroism or CD. The ellipticity of the ellipse is also a measure of the differential absorbance, known as the ellipticity θ.

$$\theta(\lambda) = 2.303(A_L - A_R)180/4\pi = 33(A_L - A_R)$$

where the absorbance A is defined as in equation (4.1) for a particular wavelength λ. The molar ellipticity is defined as

$$[\theta(\lambda)] = M\theta(\lambda)/10dc$$

where M is the molecular weight, d is the path length in centimetres and c is the concentration in

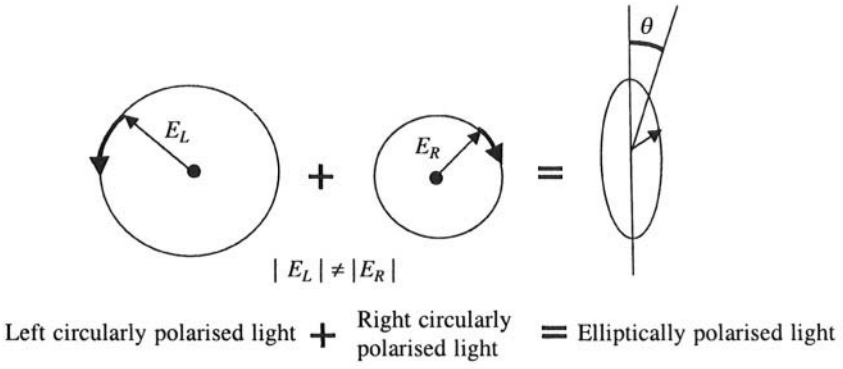


Fig. 4.4 Elliptically polarised light.

grams/ml. $[\theta(\lambda)]$ is measured in degrees. A plot of the difference in the refractive index, i.e. $(n_L - n_R)$ versus the wavelength is called the ORD curve and a plot of the difference in absorbance, i.e. $(A_L - A_R)$ versus the wavelength is called the CD curve. As seen in Figure 4.5, the two phenomena are related and together are called the Cotton effect. Both CD and ORD curves can be negative or positive and go by the names 'negative Cotton effect' and 'positive Cotton effect' respectively. Rigorously, the measurement of either CD or ORD implies the other. However, in practice, especially with the availability of appropriate instrumentation, CD is the measurement of choice.

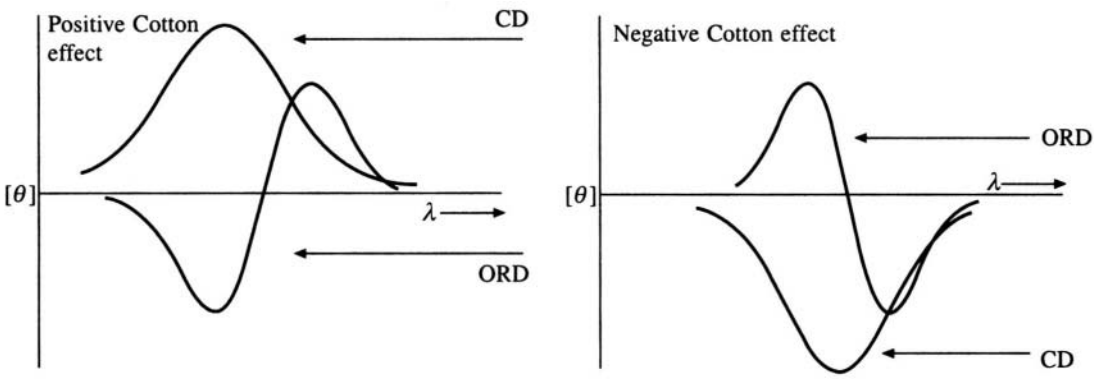


Fig. 4.5 Positive and negative Cotton effects illustrated by CD and ORD curves, respectively

The CD spectrum of a polymer is different from that of a monomer, the difference reflecting the three dimensional arrangement. CD, therefore, is immensely useful to study three-dimensional structures of biopolymers such as proteins and nucleic acids. The $\pi \rightarrow \pi^*$ transitions discussed previously dominate the spectra, and in principle, quantum mechanical calculations of the CD of the biopolymers can be performed. This would allow the interpretation of the spectrum in terms of the structure. This however is an extremely complex task and, in practice, the existing optical activity data for polypeptides, proteins and nucleic acids with known structure are used in conjunction with semi-empirical equations to analyse unknown systems.

For proteins the aim is to analyse the CD spectrum to obtain secondary structural features such as the percentage of **α-helical** regions or **β-sheet** regions or random coil regions. The protein CD spectrum is dominated by the polypeptide backbone chromophores and the effects of the side chains are usually negligible. It can be considered as a linear combination of the spectra of each of the regular secondary structural regions. One may thus write

$$[\theta(\lambda)] = k_\alpha[\theta_\alpha(\lambda)] + k_\beta[\theta_\beta(\lambda)] + k_\gamma[\theta_\gamma(\lambda)]$$

where $[\theta(\lambda)]$ is the measured ellipticity k_α, k_β and k_γ are the fractional compositions of **α-helix**, **β-sheet** and random coil regions respectively, and $[\theta_\alpha(\lambda)]$, $[\theta_\beta(\lambda)]$ and $[\theta_\gamma(\lambda)]$ are the measured CD of polypeptides in the conformation indicated. This equation can be used to estimate k_α, k_β and k_γ for an unknown protein, provided $[\theta_\alpha(\lambda)]$, $[\theta_\beta(\lambda)]$ and $[\theta_\gamma(\lambda)]$ are known. There are two methods of obtaining this information; both of them giving equivalent results when applied to unknown proteins. In the first method homopolypeptides that are known to possess one of the above secondary structures are chosen and the CD spectrum is measured for each. These spectra are used as the basis for calculating the fractional compositions. The other semi-empirical method obtains the basis values of the $[\theta_\alpha(\lambda)]$, $[\theta_\beta(\lambda)]$ and $[\theta_\gamma(\lambda)]$ from the CD spectra of proteins of known three-dimensional structure.

The second method is theoretically more attractive as the spectra will automatically include effects of length of the polypeptide on tertiary interactions.

In the case of nucleic acids, the most intense chromophores are the bases. One cannot therefore neglect the sequence and consider the backbone alone. A further contribution to the intensity arises from the stacking of the bases. Semi-empirical theoretical calculations of the CD spectrum of di-or tri-nucleotides involve treating the ellipticity due to the molecule as the sum of those due to the individual bases plus a strong contribution from the base–base stacking interactions. Thus for a trinucleotide with the bases B_1, B_2 and B_3

$$[\theta_{B_1B_2B_3}(\lambda)] = 1/3[\theta_{B_1}(\lambda)] + 1/3[\theta_{B_2}(\lambda)] + 1/3[\theta_{B_3}(\lambda)] + I_{B_1B_2} + I_{B_2B_3} + I_{B_3B_1}$$

where $I_{B_1B_2}$ . . .etc., are due to the base-base interactions. Note that the term $I_{B_3B_1}$ represents a 'next nearest neighbour' interaction rather than the 'nearest neighbour' interaction of the other two terms. Normally one needs to consider only the 'nearest neighbour' interactions. This approach of building up the CD spectrum of a polynucleotide from the spectrum of the monomer can be extended quite successfully to longer chains, though the number of contributions then increases. For DNA the base-paired double-helical structure must also be taken into account. Accurate determinations of the average structures of several different synthetic polynucleotides with several different structures have been made. A dramatic example of the use of CD was when an inversion of the high salt spectrum of poly d(CG), lead to the prediction of a novel left-handed double helical structure.

4.4 Fluorescence Spectroscopy

Fluorescence is a phenomenon by which light is absorbed by a system at one wavelength and emitted by it at a different, and longer, wavelength. Measurements of fluorescence give information about molecular conformation and dynamics and form an important part of the techniques used in biophysics. The basic principle of fluorescence can be understood in terms of the electronic energy levels in one of which, according to quantum mechanics, the molecule is usually found (Figure 4.6). Each of these electronic levels themselves has a so-called 'fine structure' due to the vibrational motion of the molecules, i.e. each electronic level is not a single energy value but consists of a band of closely spaced energies which arise due to the different vibrational modes. The molecule usually exists in the ground state. When a beam of light (visible or UV) falls on it, some of the energy is absorbed and electrons jump up to the excited state where they can go to any of the vibrational levels. So far the experiment is the same as UV/visible absorption spectroscopy. However, now the electron in the upper electronic level can lose some of its energy in many ways. One of these is by going to a lower vibrational level within the same electronic level. When now the electron jumps back to the ground state, it does so by emitting radiation, but the radiation is of less energy and therefore longer wavelength than the incident light. The molecule is said to fluoresce. A study of the intensity of the fluorescent radiation as a function of the wavelength is known as fluorescence spectroscopy. Fluorescent radiation is spontaneously emitted radiation as opposed to stimulated emission. It is governed by a rate constant k_F

$$k_F = A_{ba} = 1/\tau_F$$

where τ_F is called the radiative lifetime of the excited electronic state (b) of the molecule. A_{ba} is the Einstein coefficient of spontaneous transition from state (b) to state (a), which is the lower, ground state. A_{ba} can be evaluated in terms of the frequency of the incident light and the strength of electric moment dipole corresponding to the states (a) and (b). The molecule can also go from the excited state

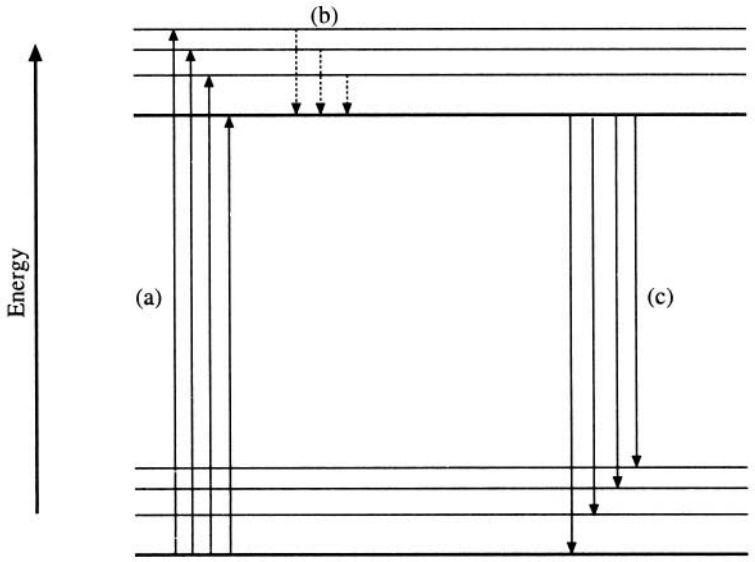


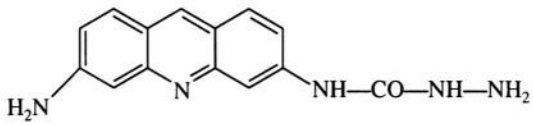
Fig. 4.6 Electronic energy levels in a molecule, indicating the fine structure, and possible transitions. (a) Transitions from ground to excited state. (b) Non-radiative loss of energy. (c) Radiative transitions to lower state

to a lower state by several other mechanisms that do not involve emission of radiation. For example, the excitation energy may be lost through collisions with the solvent molecules or by increasing the vibration of the molecules. The rate at which energy is lost through this process is called k_{ic} for 'internal conversion'. Another way in which the energy can be dissipated is by a phenomenon called quenching, resulting from the formation of complexes with solute molecules. The rate constant governing this process is called k_Q. A third process is when the excited state is converted to another, excited state of a lower energy though the transition that may be normally disallowed by quantum mechanics. The new excited state then dissipates its energy by internal conversion or by emission at a much longer wavelength than the fluorescence. The rate constant for this process is designated as k_{is} for 'intersystem crossing'. Thus the actual proportion of the absorbed incident light which is reemitted as fluorescent radiation is small and is called the quantum yield. It is given by the expression

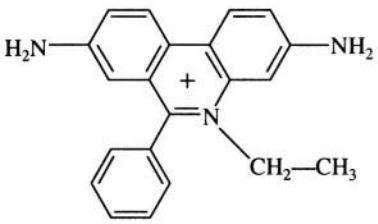
$$\phi_F = k_F/(k_F + K_{ic} + k_Q + k_{ia})$$

Chromophores or UV/visible absorption groups in biological molecules like proteins and nucleic acids have rather low quantum yields and very short radioactive lifetimes. For example, the most intensely absorbing group among the constituents of proteins is the tryptophan side chain. Its absorption maximum is at 280 nm while its fluorescence emission maximum is at 348 nm. It has $\phi_F = 0.20$ and $\tau_F = 2.6$ nanoseconds. The tyrosine side chain absorbs at 274 nm, emits at 303 nm, $\phi_F = 0.14$, $\tau_F = 3.6$ nanoseconds. Phenylalanine absorbs at 257 nm, emits at 282 nm, has a very poor quantum yield of 0.04 and a lifetime of 6.4 nanoseconds. Other side chains have yields much lower than this. In nucleic acids, the bases have yields of the order of 10^{-4} and lifetimes less than 0.02 nanoseconds. It may be noticed that the strongest fluorescent chromophores are rigid planar moieties, since in such groups, the energy dissipation by vibrational and rotational motion is much less. The flexibility of the nucleic acid backbones and of the other side chains and backbones of proteins is responsible for the fact that virtually no fluorescence is detected from them.

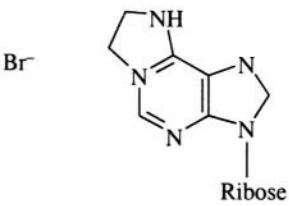
Nevertheless, since fluorescence is extremely sensitive to the environment of the chromophore, it is an effective technique to study conformational changes and ligand binding. The studies are enhanced by the use of fluorescent probes (Figure 4.7). Some of these have the property that their fluorescence is strongly quenched in an aqueous solvent, but is significantly elevated in a rigid or non-polar environment. The binding of ethidium bromide to double-stranded DNA or RNA by intercalation between the base pairs can therefore be followed very easily by fluorescence spectroscopy and this ligand is used as a probe to detect the double helical structure of nucleic acids. The use of extrinsic chromophores, as these ligands are called, helps also in studying phenomena related to proteins such as the binding of the substrate, the position of various residues in the protein, etc.



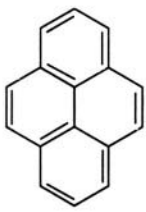
Proflavine monosemicarbazide



Ethidium Bromide



Ethenoadenosine



Pyrene

Fig. 4.7 Some common fluorescent probes

If polarised light is used as the incident beam, a measurement of the polarisation state of the emitted fluorescence yields information on the mobility of the chromophores. This is because as mobility increases, the plane-polarised light becomes more and more circularly polarised. Applications of polarisation fluorescence spectroscopy include the study of the helix-coil transition in proteins and the orientation of DNA in chromosomes.

A word about the experimental set up (Figure 4.8). Unlike as in absorption spectroscopy or in CD/ ORD, the source of radiation, the sample and the detector are not placed along the same straight line. This is because fluorescence emission occurs in all directions and placing the detector at right angles increases the sensitivity by cutting out the direct beam.

4.5 Infrared Spectroscopy

Infrared radiation is heat radiation and term 'infrared' extends from 10^{14} to 10^{11} Hz on the electromagnetic spectrum. The energy of infrared radiation corresponds to energy differences between different vibrational modes in molecules. Infrared spectroscopy is therefore a probe of the vibrational motion of the molecules. Not all types of vibrations can be detected by IR spectroscopy. Since the technique involves the interaction of electromagnetic radiation with the molecules only those vibrational transitions which are accompanied by a change in the dipole moment can be detected. This means that diatomic molecules such as H_2, O_2, N_2, do not exhibit infrared absorption and the vibration states of such molecules cannot be examined through IR spectroscopy. This is because the dipole moment is a

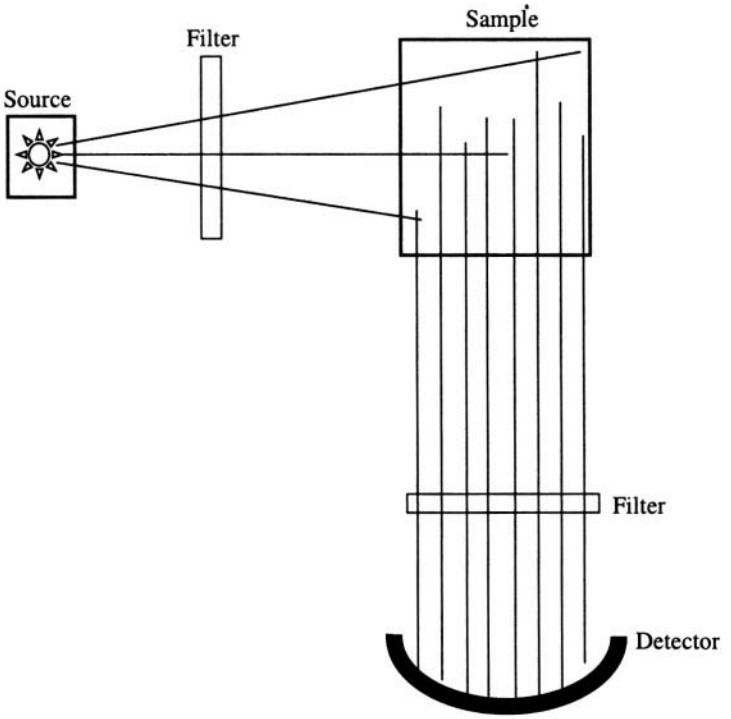


Fig. 4.8 A schematic sketch of the experimental setup used in fluorescence spectroscopy

measure of the separation of the centres of total positive charge and the total negative charge of the molecule. Diatomic molecules can only undergo symmetric vibrations, which will not change the dipole moment. For similar reasons, certain types of vibrations in particular molecules are IR-active, while others are IR-inactive. For example, in carbon dioxide, a symmetric stretching vibration of the oxygen-carbon bonds (Figure 4.9) is IR-inactive, while asymmetric stretching or bending vibrations are IR-active.

Figure 4.10 shows, as an example, the IR spectrum of 6-benzoyl adenine. It may be seen that vibrational absorption spectra consist of bands rather than lines. The spread in the energies is usually due to conformational differences in the molecules and free rotation about the bonds. Still, the spread is not as large as for UV absorption. IR spectral bands are narrower in the solid state and at low temperatures. IR spectra are plotted as a function of wave number, i.e. the reciprocal of wavelength

$$\text{Wave number} = 1/\lambda \text{ cm}^{-1}$$

Sometimes the percent transmittance rather than percent absorbance is plotted. The two quantities are complementary, i.e.

$$\text{Transmittance} = (1 - \text{Absorbance})$$

Except for the simplest molecules, infrared spectra are often characterised by many absorption bands that are broad and closely overlapping. Therefore group vibrations have been extensively used to determine whether particular chemical groups are present and to chemically characterise the samples. Groups that absorb in those regions of the spectrum, which do not have too many other absorption bands, are selected. One such region for biological molecules lies between 1800 and $2300\ \text{cm}^{-1}$. The

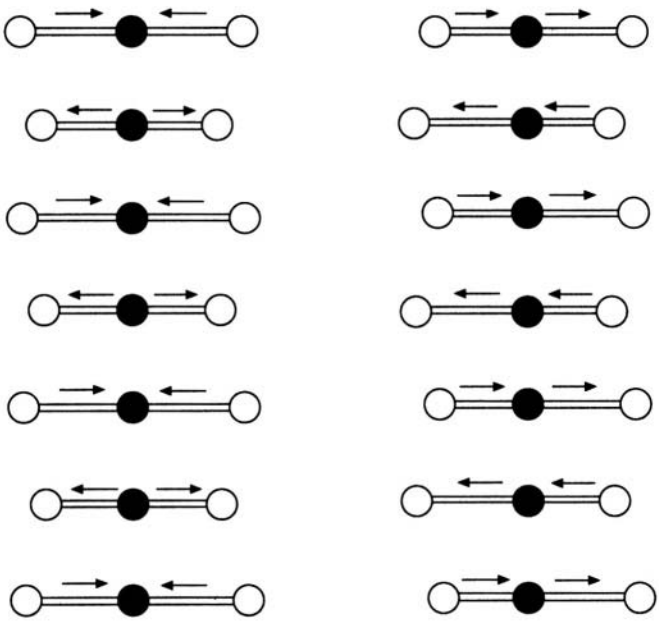


Fig. 4.9 The IR (a) inactive and (b) active stretching modes in carbon dioxide

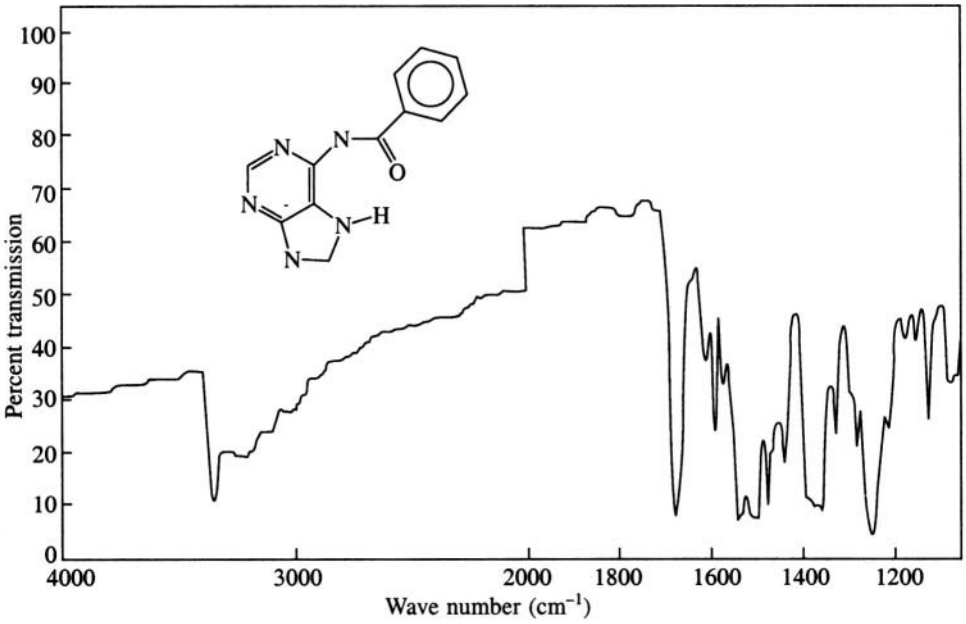


Fig. 4.10 Infrared spectrum of 6-benzoyl adenine

lower **400–1800 cm^{-1}** region of the spectrum has many overlapping bands, and is called the 'finger print' region. For relatively simple structures, this region is used for identification. Table 4.2 gives a list of group absorption bands.

Table 4.2 Characteristic IR bands in polypeptide chains

Bond	Mode of vibration	Wave number (cm^{-1})
C–H	Stretch	2700-3300
C–H	Bend	1300 – 1800
O–H	Stretch	3030 – 3700
O–H	Bend	1200 – 1800
N–H	Stretch	3000 – 3700
N–H	Bend	1500 – 1700
C=O Amide I hydrogen bond	Stretch	1630 – 1660
C=O Amide I non-hydrogen bond	Stretch	1680 – 1700
C–N–H Amide II hydrogen bond	Stretch	1520 – 1550
C–N–H Amide II non-hydrogen bond	Stretch	< 1520

The amide I and the amide II bands are the principal infrared absorption bands of the peptide group in proteins. As may be seen from the table, both bands absorb in different regions according to whether they do or do not participate in hydrogen bonding interactions. Hydrogen bonds affect the vibration frequencies of the participating atoms. The presence of H-bonds is an indication of the polypeptide chains assuming regular secondary structures. In fact both amide I and amide II bands absorb in slightly different regions if they are in **α-helices** than if they are in **β-sheets.** This helps in the study of the conformation of biopolymers.

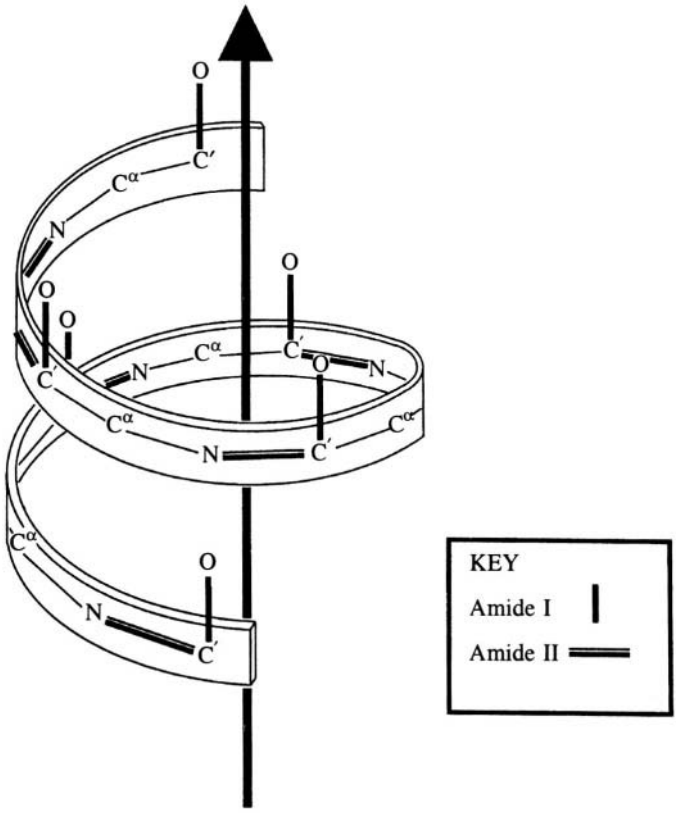
Carbon-monoxide stretching frequencies serve as a marker to study the conformational changes on the binding of this ligand to haemoglobin, the oxygen carrier protein. Studies on haemoglobin have also used sulphydryl group frequencies of the cysteine group in the protein. The S–H absorption frequency centres around **2500 cm^{-1}** and is slightly different for the three different cysteine groups in haemoglobin. The utility of this group as a conformational marker lies in the fact that its absorption band lies in a region away from that of water, thus making it possible to study the protein in aqueous solutions. Other absorption bands would be swamped by water frequencies. Dichroic effects on oriented polypeptides can be observed using polarised infrared radiation. For example, if the polypeptide is a **α-helix,** the N-H and C=O groups and the hydrogen bonds between them are oriented along the helix axis. The bonds that make up the amide II band are perpendicular to the helix axis (Figure 4.11). In **β** sheets the situation is just the opposite, with the N-H and amide bands being perpendicular to the chain direction, and the amide II bands being parallel to it. The absorption of polarised IR radiation consequently depends on the orientation of the sample relative to the beam. The spectrum of the actual polymer is however much more complex than the above discussion suggests. All the same, infrared dichroism of oriented polypeptide films is an established technique to study conformation.

Infrared biospectroscopy has in recent times benefited a great deal from improvements in measurement techniques. These improvements have resulted in large increases in the signal-to-noise ratio. One of the improvements was the change from using a continuous scanning mode to sweep across the

frequencies, to an interferometric, Fourier transform spectroscopic mode. The principle behind this technique is to irradiate the sample by a sharp pulse of IR radiation that contains all the relevant frequencies and then detect the response as a function of time. A Fourier transform of this detected signal yields the conventional absorption spectrum. [The Fourier Transform technique is explained in greater detail in the chapter on NMR spectroscopy.]

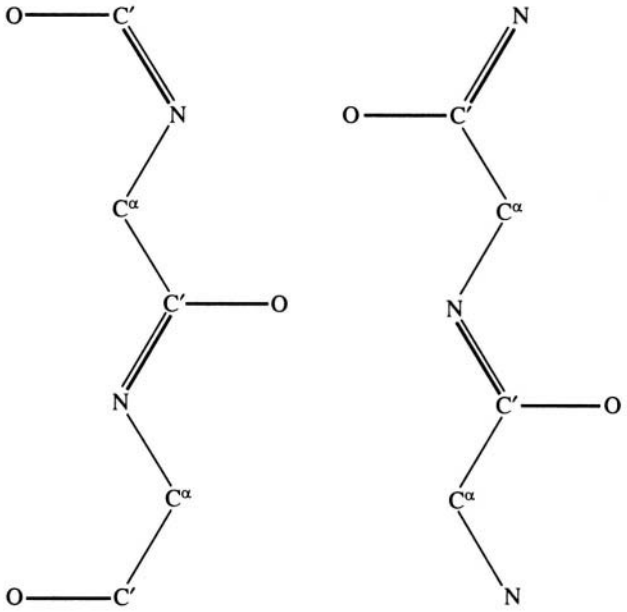
4.6 Raman Spectroscopy

The Raman effect was discovered by C.V. Raman and his associates in 1928. It arises when a beam of light (electromagnetic radiation in the visible region) is allowed to fall on a transparent substance such as a solution of a biological molecule. Most of the incident light will be transmitted through the sample unaltered, but a small portion is scattered at angles different from the incident direction. Again much of this scattered light is at the same frequency as the incident light. This scattering is called Rayleigh scattering. (The intensity of Rayleigh scattering is an inverse function of wavelength. Therefore a polychromatic incident beam will have a larger proportion of its smaller wavelength components scattered than its longer wavelength components. This is why the sky is blue). Apart



Left-handed α-helical polypeptide chain
(a)

Fig. 4.11 (*Contd*)



Polypeptide chains in anti-parallel β sheet arrangement
(b)

Fig. 4.11 A schematic representation of the orientation of the bonds which produce the amide I and the amide II bands in polypeptide chains. (a) The α-helical conformation is a left-handed α helix, but the orientation of the bonds in a right-handed helix is the same. (b) The β-sheet conformation is an anti-parallel sheet, but the conformation of the bonds in a parallel sheet is approximately the same

from Rayleigh scattering, a small portion of light is scattered with an altered frequency. This phenomenon is called Raman scattering. The shift in the frequency corresponds to the vibrational energy states of the molecule. Raman scattering is thus a probe of the vibrational states of the molecule, just like IR spectroscopy, except that the incident radiation is not infrared, but visible light and what is measured is not the absorbance but the frequencies of the scattered light.

The oscillating electric field of the incident light induces a dipole moment in the sample, different from any permanent dipole moment that may be present. As already seen, the permanent dipole moment gives rise to IR absorption. The induced dipole moment gives rise to Raman scattering. The strength of the induced dipole moment depends on the polarisability of the sample. Both classical mechanics and quantum mechanics show that if a particular transition from one vibration mode to another is accompanied by a change in the polarisability, the incident light will be absorbed and re-emitted with an altered frequency, leading to the Raman spectrum. The Raman effect is thus a coupling of the electronic and vibrational states of the molecule. Since, before the transition, the molecule could exist either in the higher or lower vibrational state, the scattered light could be either of lower or higher energy than the incident light. The light of lower frequency is called the Stokes spectrum or the Stokes lines. The longer frequency emissions are called the anti-Stokes lines, and are usually quite weak as compared to the Stokes lines. If it appears paradoxical that the emitted light should be of a higher frequency than the incident light, the Raman effect may be considered as a 'two-photon' phenomenon. Two photons are absorbed, the energy is exchanged and then re-emitted so that

one has lesser energy and the other higher energy than the incident photons. Usually only the Stokes spectrum is plotted, with intensity along the y axis and wave number increasing along the negative x direction (Figure 4.12).

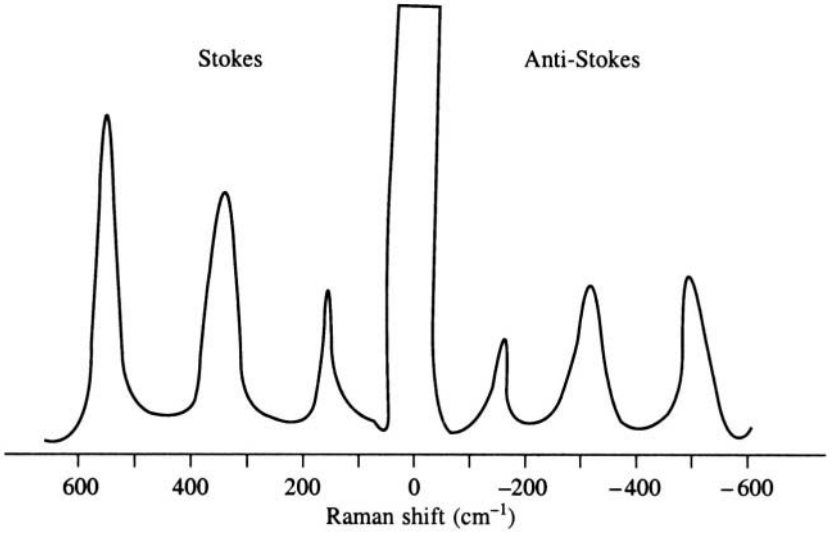


Fig. 4.12 Schematic of a Raman spectrum.

Since only a small portion of the incident energy is re-emitted as Raman scattering, the source has to be quite intense in order that the effect is observed. Also the concentration of the sample needs to be high. For these reasons, it was only after the introduction of lasers that Raman spectroscopy of biomolecules was possible. Though Raman spectroscopy, like IR, also probes vibrational modes of the molecules, there are differences and advantages. In the first place, since the mechanism of the Raman effect is different from that of IR absorption, several vibrational transitions that do not change the dipole moment but change the polarisability, can be detected only by Raman spectroscopy. Secondly, and this is of great advantage, water does not interfere very much in the Raman spectrum. This makes possible the study of aqueous solutions of biomolecules. Other advantages lie in the sensitivity of the Raman effect to conformational changes and in the possibility of using various types of samples, including powders, liquids, films, etc.

In the application of Raman spectroscopy to proteins the amide bands are again used to probe the structure. The ones that are most sensitive are the amide I band **(1645–1680 cm^{-1})** which is predominantly a C = O stretching mode, the amide III mode **(1235–1300 cm^{-1})** which results from a mixture of C–N stretching and N–H in plane bending, and the C–C stretching at **900-1000 cm^{-1}.** The amide II band between 1230 and **1310 cm^{-1}** is also used to distinguish between **α-helix, β-sheet** and random coil conformations of biomolecules. Raman spectra of entire proteins have been obtained and as expected they are quite complicated. The peaks however are quite sharp and in general one may consider the spectrum of such a molecule as the sum of the spectra of the constituents, in this case, the amino acids. This, of course, is especially true in proteins such as lysozyme, where a large portion of the chain is in the random coil conformation. Raman spectroscopy has been used effectively to study thermal unfolding. In ribonuclease, a decrease in the intensity of the amide I and amide III lines corresponding to the **α-helical** and **β-sheet** frequencies was seen as the temperature was increased. This gave an indication of the unfolding of the protein.

The application of Raman spectroscopy to nucleic acids has been somewhat more extensive, since these polymers are constituted from only four monomer units, the bases, unlike proteins where there are 20 amino acids. The Raman spectra of nucleic acids has two distinct classes of lines; those arising from the bases, and those due to the sugar-phosphate backbone. The phosphate group has two types of P–O symmetric stretching modes. One arises from the phosphoester bonds and is around $810\ cm^{-1}$. This band is very sensitive to the conformation and on the basis of this band alone, a change of the double-helical conformation from *A* type to the *B* type may be monitored. The other symmetric P–O stretching arises from the bonds to the charged pendant oxygens. This occurs around $1100\ cm^{-1}$ and being relatively insensitive to conformational change, serves as an internal marker. Certain Raman bands also give information about base pairing. The Raman hyperchromic and hypochromic effects, which arise from base-base interactions, are probes of the helical ordering of the polynucleotide chains.

Raman spectra of liquids and membranes are used to follow order-disorder transitions, through intermediate liquid crystalline phases. Some of the bands which are monitored are the C–H stretching region between 2700 and $3100\ cm^{-1}$, the C = C vibrations at $1650\text{-}1700\ cm^{-1}$ and the skeletal C–C stretching at $1050\text{-}1150\ cm^{-1}$. Raman spectra of intact organelles or organisms such as bacteriophages, chloroplasts, and ribosomes have also yielded significant structural information.

With the availability of tunable lasers, the technique of resonance Raman spectroscopy has become feasible. In this technique the incident laser beam is tuned to be in resonance with electronic transitions in the chromophore. Sample concentrations as low as 10^{-5} M may be used. A further advantage is that resonance Raman scattering yields information only about the immediate environment of the chromophore. This allows experiments where the chromophore is diffused into the active site of a protein, whose structure is then probed by Raman scattering. Heme groups and retinols are especially useful in this technique.

4.7 Electron Spin Resonance

Electrons possess charge and spin. Thus they have a magnetic moment. For a free electron, the magnetic moment is given in terms of the Bohr magneton, similar to the nuclear magneton. When such an electron is placed in a magnetic field *H,* the interaction of the field with the magnetic moment gives rise to a Larmor precession, the frequency of which is proportional to βH, i.e.

$$\nu = g\beta H/h$$

where g is a proportionality constant called the splitting factor or the 'g' factor and h is Planck's constant. The value of g is 2.00232 for free electrons. The precession frequency ν is usually of the order of gigahertz (GHz, 10^9 Hz) as compared to nuclear precession frequencies of MHz. Just as in NMR, the free electrons absorb energy from an electromagnetic field oscillating exactly at their Larmor frequency. This absorption, when the resonance condition is satisfied, can be detected and is called electron spin resonance or electron paramagnetic resonance. The utility of this phenomenon in chemistry and biology comes from the variation that the splitting factor undergoes in different chemical environments.

When the electrons are not free, but bound to some nucleus, ESR can be detected only for an unpaired electron. According to Pauli's exclusion principle, pairs of electrons will always be arranged so that the directions of the two spins are antiparallel. This implies that the interactions of the magnetic moments of the two electrons with the field will cancel each other. Only for an unpaired electron will there be a detectable interaction. The magnetic moment of an unpaired electron is about 700 times that of protons. Hence the sensitivity of an ESR experiment is higher than that of an NMR

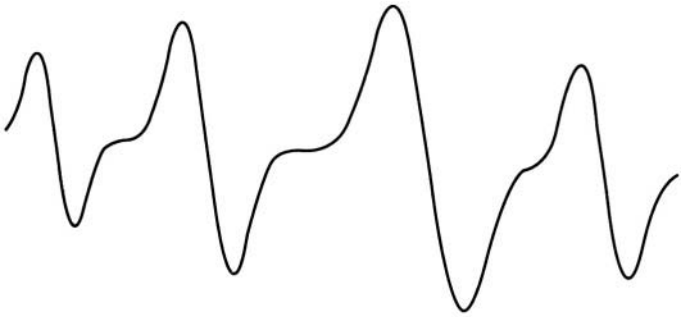


Fig. 4.13 Schematic of an ESR spectrum

experiment and the quantity of the sample required to detect the signal is comparatively small. In many ways the ESR experiment is very similar to the NMR experiment and factors such as the T_1 and T_2 relaxation times play a key role. Since ESR absorption lines are broad compared to the NMR spectrum, it is usual to plot and analyse the first derivative of the absorption spectrum rather than the spectrum itself (Figure 4.13). The role of the chemical shift in NMR is played by the g factor in ESR. The values of g for some organic radicals are shown in Figure 4.14. Note that only the oxygen and the nitrogen radicals have large differences in the value of g from that of a free electron. The magnetic moment of the electron also interacts with the nuclear magnetic moments in the neighbourhood. This interaction leads to a splitting of the absorption lines known as the hyperfine structure or the hyperfine splitting. The origin of this splitting may be understood as follows. Consider an unpaired electron in the external magnetic field H. The field experienced by this electron is a combination of local fields due to surrounding nuclei and the external field. Thus the effective field on the electron is

Organic radical	'g' value
—H	2.00226
—N(=O)—O	1.999
—Cl(=O)—O	2.01

Fig. 4.14 Typical value of the splitting factor or g factor for some common organic radicals

$$H_{\text{eff}} = H + \Delta H_{\text{loc}}$$

where ΔH is the (small) local field. In the case where the electron can be considered to be localised around just one nucleus, of spin quantum number I, ΔH_{loc} will assume $2I+1$ values along the direction of the external field. Resonance therefore will occur at the following $2I + 1$ values of H_{res}:

$$H_{\text{res}} = H^{\text{O}}_{\text{res}} - am_I, m_I = -I, -I + 1, \ldots, I - 1, I$$

where a is a constant such that am_I is the value of the local field and H^0_{res} is the field strength at which resonance would occur if $a = 0$. Figure 4. 15 shows the situation for a hydrogen atom with its single electron and the single proton in the nucleus, $I = 1/2$. The dashed lines show the case when a = 0. The solid lines are the two transitions that actually occur, giving rise to the indicated hyperfine splitting. In general, if the spin quantum number of the nucleus is I, the ESR spectrum is split into $2I + 1$ lines.

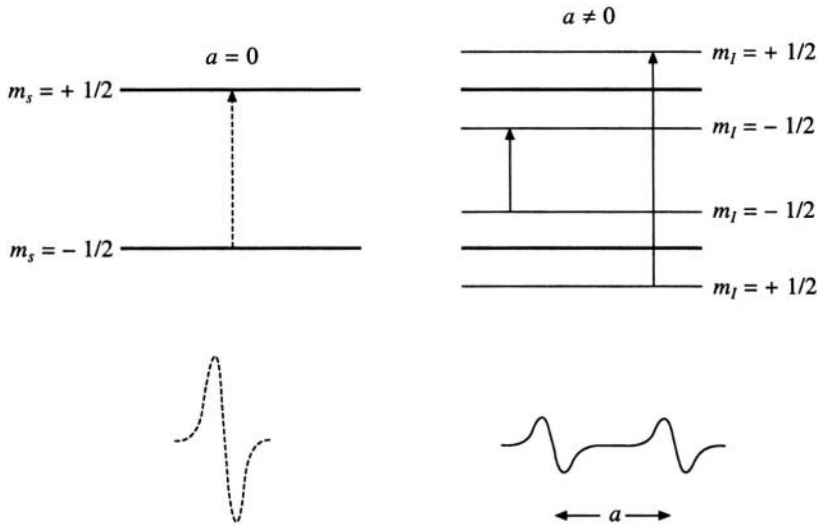


Fig. 4.15 The hyperfine splitting in hydrogen

ESR spectroscopy yields useful information in biological systems. The binding of paramagnetic metal ions such as Cu^{2+}, Mn^{2+}, etc., (i.e. those with an unpaired electron) to proteins or other molecules can be studied by the splitting that the binding induces in the absorption lines. Both the common isotopes of nitrogen, i.e. N^{14} as well as N^{15} will lead to a splitting of the absorption line of Cu^{2+} when the metal ion binds. The resonance frequency of a paramagnetic species also depends on the orientation of the spin with respect to the external magnetic field. In a liquid, the tumbling motion will average out all the orientations. But in a crystalline solid, the ESR spectrum contains information regarding the orientation. This property was used, for example, in determining the angle of the heme group, containing paramagnetic iron, with respect to the crystal axes. Also, the orientation effect leads to an inverse correlation between the tumbling motion of the molecule and the width of the absorption line – the wider the line, the slower the tumbling. This has helped in the study of fluidity of membranes and the dynamics of proteins.

In the case of molecules which do not have intrinsic paramagnetic atoms (as in most biological macromolecules), ESR studies are made by attaching one or more organic or other radicals. Such paramagnetic groups are known as spin-labels and act as a reporter or probe. For most studies spin-labels containing the nitroxide group are used. The ESR spectrum of the spin label can then be analysed in terms of the width of absorption line and the hyperfine splitting, to yield information about the conformation and dynamics of the biological molecules. This technique is especially useful in the study of membrane bilayers.

5

Light Microscopy

5.1 Introduction

Many of the interesting features of biological systems are too small to be seen with the naked eye. When molecular detail is not required, the light microscope is an ideal, and hence essential, instrument for a biologist. Electron microscopy can also be used to elicit highly detailed information, but suffers from several limitations, including complex operating procedures. Light microscopy is easy to use and inexpensive. An additional, crucial advantage it has over the electron microscope is its ability to focus at different levels within a three-dimensional object. In addition, the new methods and applications, which are even now being constantly developed, have helped to retain the importance of the light microscope in biological studies, in spite of the availability of advanced instruments of all kinds based on a variety of other physical principles.

Light microscopes are built up from a stack of lenses, usually made of glass and in this chapter, the optical principles which govern their arrangement are discussed first. A discussion of the resolving power of microscopes follows. Finally we discuss the newly developed methods in microscopy.

5.2 Elementary Geometrical Optics

Optical lenses can be either concave or convex. Convex lenses (Figure 5.1 (a)) are also called converging lenses. Concave lenses (Figure 5.1(b)) are known as diverging lenses. Each lens has two focal points as shown. In the figure the focal lengths, defined by OF_0 and OF_1 are equal. This is not always so and depends on the relative refractive indices of the lens and the medium on either side of it. A thin concave lens always gives a minified (i.e. made smaller) virtual image of any object viewed through it (Figure 5.2). Virtual images are images that cannot be captured on a screen or on photographic film, except with the help of additional lenses. Concave lenses are thus not normally used in microscopes expect in combination with other lenses. A real image, in contrast, can be captured on a screen. Convex or converging lenses may form real or virtual images, depending on the positions of the

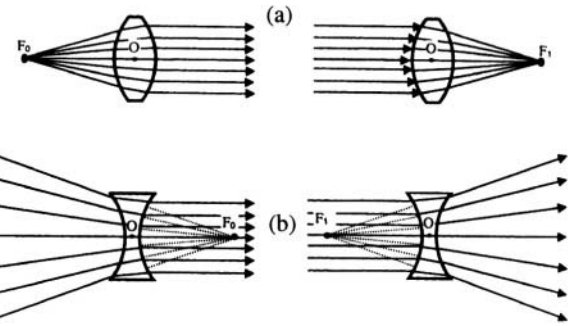


Fig. 5.1 (a) Convex and (b) Concave lenses.

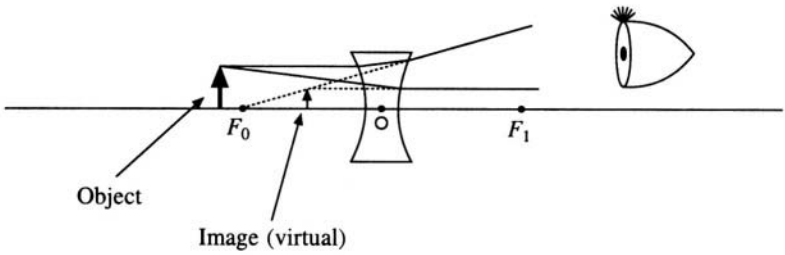


Fig. 5.2 Image (virtual) formation by a concave lens

object with respect to the focal points of the lens. Also the images may be minified, magnified, or of the same size. Figure 5.3 illustrates the various possibilities. When the object is at a distance greater than twice the focal length from the lens, a real, inverted, minified image occurs at a point between one and two times the focal length on the other side of the lens (Figure 5.3(a)). When the object is at a point exactly twice the focal length, the image is real, inverted and the same size (Figure 5.3(b)). When the object is at a distance between one and two times the focal length, the image is real, inverted and magnified, and occurs at a distance greater than twice the focal length (Figure 5.3(c)). When the object is at the focal point, the image occurs at infinity (Figure 5.3(d)). The basic compound microscope uses two convex lenses in the arrangement shown in Figure 5.4. The lens closest to the object is called the 'objective'. The lens closest to the eye is called the 'eyepiece'. The microscope is adjusted so that the object (a) is at a distance between one and two times the focal length. The image (b) formed is therefore real, magnified and inverted. This real image acts as the object for the second lens, i.e. the eyepiece. The eyepiece magnifies this intermediate image, and produces a large virtual image (c) which then becomes the image for the eye or for the camera lens. The final real image (d) is formed on the retina of the eye or the photo film. The relative positions of the eyepiece and the objective are kept constant and focusing the microscope requires movement of both as a single element. The overall magnification of the instrument is the product of the magnifications of the objective and the eyepiece. In most microscopes these are changeable and various combinations of the two lenses can be used to obtain the required magnification. In some microscopes, the magnification can be 'zoomed' up and down in a continuously variable fashion. Lenses with aspherical surfaces are also available, to reduce optical aberrations. In fact, even the simplest of compound microscopes available today will not be as simple as the one shown in the figure. The objective is a combination

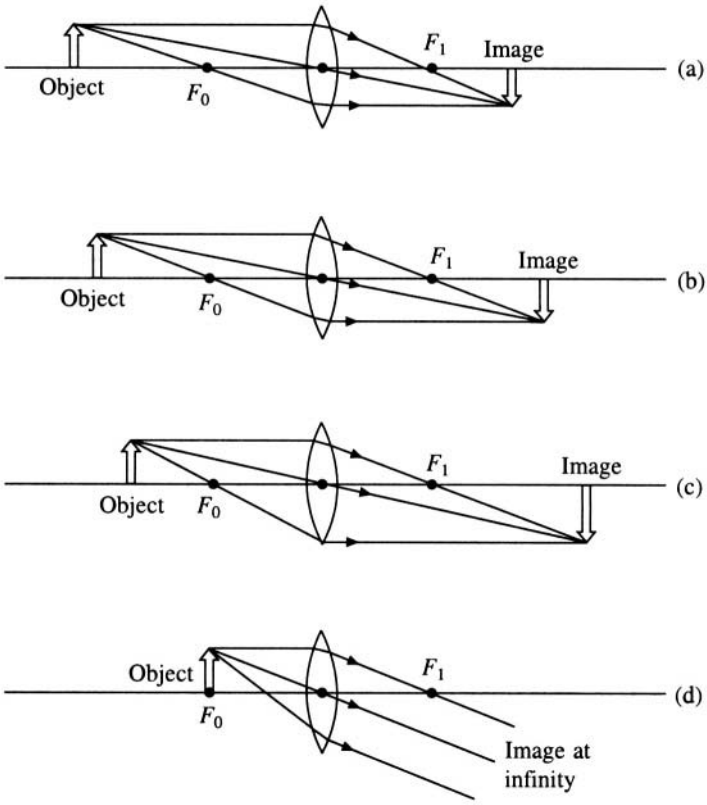


Fig. 5.3 Image formation by a convex lens

of convex and concave lenses to produce the desired magnification and eliminate distortions. Similarly the eyepiece is built up of many lenses. Nevertheless, the principles underlying the magnifying systems remain as illustrated here.

5.3 The Limits of Resolution

The resolution of a microscope is its ability to distinguish closely spaced objects. Heisenberg's Uncertainity principle (see Chapter I) states the theoretical limits to the amount of magnification that an object can be subjected to and this figure is far, far greater than the magnifying power of any light microscope. For the magnification to make sense, however, the final image should make visible the individual features. The smallest distance at which two point like objects can be placed and still appear as two distinct objects governs the actual, effective magnifying power of the microscope. The smaller this distance, the higher the resolution of the microscope and, therefore, the higher the magnification possible. The resolution limit of an instrument can be defined on the basis of Rayleigh's criterion, which, in turn, comes from an analysis of the diffraction of light. Light waves coming from different points of the object form the final image after travelling through slightly different distances. As explained in the chapter on X-ray diffraction, the waves can interfere constructively or destructively to form a diffraction pattern. Figure 5.5 shows the image, i.e. the diffraction pattern that is formed by a small pinhole. A central bright image is surrounded by concentric rings, which rapidly become fainter. The central image is called the first maximum, and the dark ring immediately surrounding it

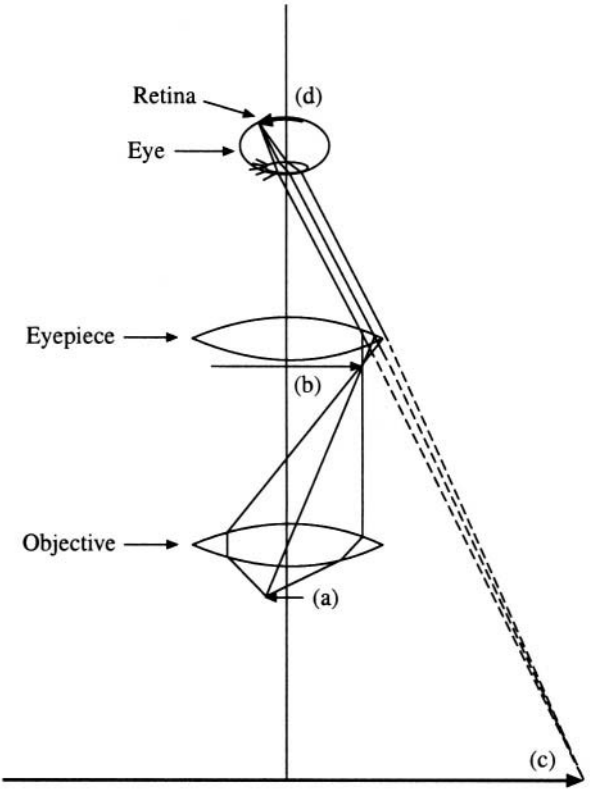


Fig. 5.4 Schematic diagram of a compound microscope

is the first minimum. The diffraction pattern due to two pinholes close to each other will overlap to different extents depending on the distance between the holes. Rayleigh suggested that two such objects could be considered as resolved, if in the diffraction pattern, the first maximum of one object overlapped with the first minimum of the other (Figure 5.6). Rayleigh's criterion is especially useful in deciding the resolution limit when dealing with self-luminous objects, such as for example, stars viewed through a telescope.

Another obvious limit to the magnifying power of a microscope is the wavelength of the light used for illumination. If the spacing between the objects is less then about 1/2 the wavelength of the light, then the images of the two objects will almost exactly superimpose on each other and they cannot be resolved. For the usual type of light used in microscopes this limit is about 250 nm. Considering the fact that the eye (or the photographic film) imposes its own limits, the total magnification that is usually possible with light

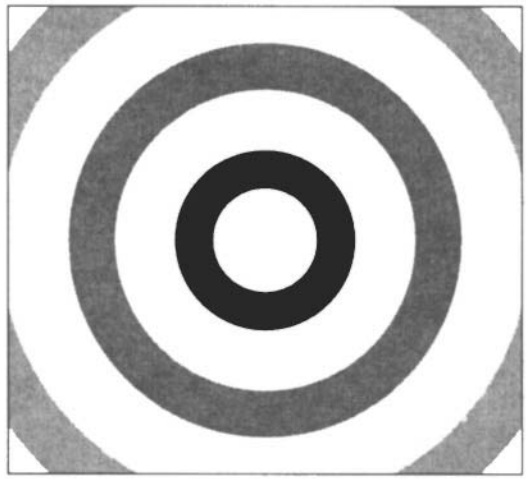


Fig. 5.5 Diffraction pattern from a pinhole

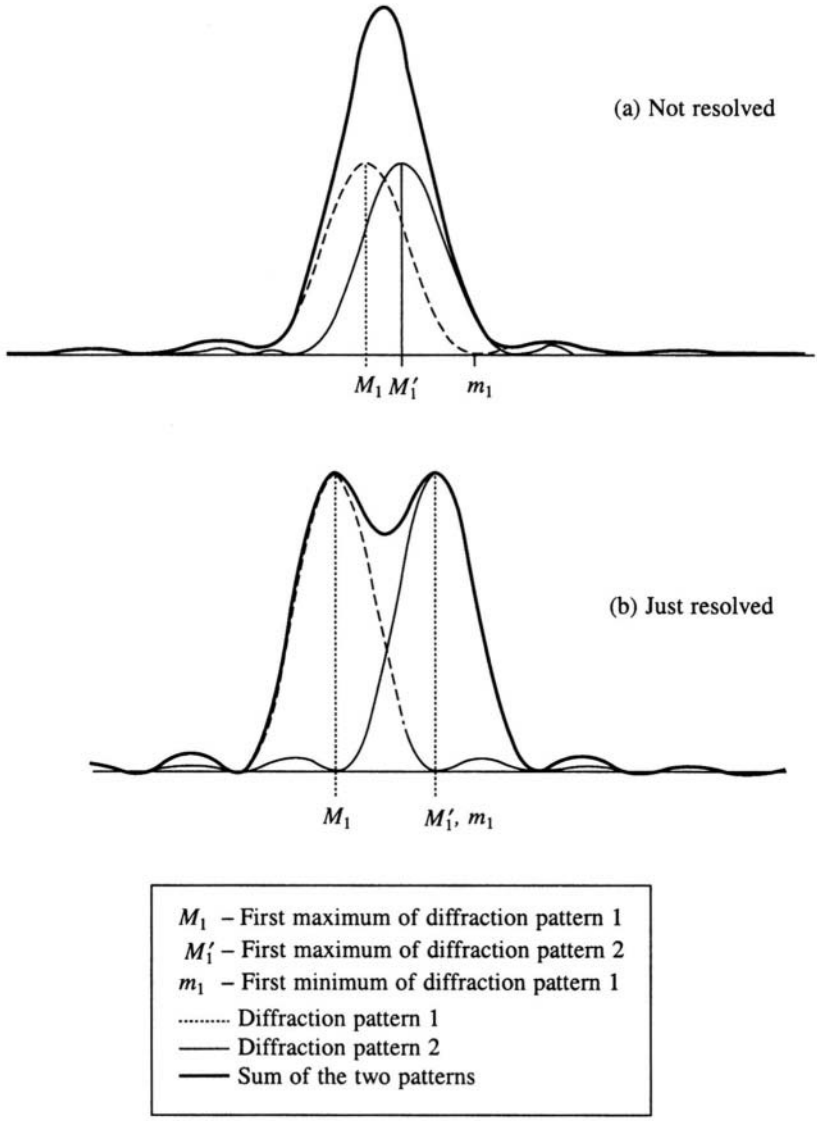

(a) Not resolved
M_1 M_1' m_1
(b) Just resolved
M_1
M_1', m_1
M_1 – First maximum of diffraction pattern 1
M_1' – First maximum of diffraction pattern 2
m_1 – First minimum of diffraction pattern 1
Diffraction pattern 1
Diffraction pattern 2
Sum of the two patterns

Figure 5.6 Rayleigh's criterion for the resolution of two objects

microscopes using ordinary light is about ×800 (×800 or 800× means 800 times). While the Rayleigh criterion is simple, there are certain disadvantages when applied to images formed by microscopes. Apart from being strictly applicable only to self-luminous objects, there is also the problem of analysing diffraction from a complex object. The approach of Abbe is probably more suited to microscopes. Abbe realized that the objective lens in a microscope acted essentially as a Fourier summation device. Fourier showed that any regular periodic function could be written as a sum of simultaneously varying waves of different wavelength and amplitude. When light falls on an object, the effect of diffraction is to split the wave into Fourier components. A summation of all these components will give back the original image and this task is performed by the objective lens.

However, owing to the large angles at which most of the components are scattered, the objective usually is able to capture only the low order components. This leads to a fuzzy representation of the objects. The angles at which the various components are scattered also depends on the spacing of the features in the object—the smaller the spacing, the larger the angle. Thus features which are very close together are not well resolved since fewer of the Fourier components of the diffraction pattern from these features are used by the object lens in the Fourier summation. Abbe suggested that a pair of features on the object be considered as well resolved if both the zeroth and the first orders of the Fourier components reach the objective. This criterion leads to essentially the same result as the Rayleigh criterion, for the resolving power of a microscope and the highest meaningful magnification. It has the additional advantage of providing a rigorous understanding of the similarities between the microscope lens system and the mathematical operations of Fourier transformation.

5.4 Different Types of Microscopy

5.4.1 Bright field microscopy

Bright field microscopy is obtained in the normal method of illuminating the object, when thin, more or less transparent samples (such as cells or bacteria) are being viewed. The illumination is usually as illustrated in Figure 5.4. The light is transmitted through the sample and the image is formed by the absorption of this light. Thus the image will appear darker than the background which is the bright field. In other words, the image is formed due to the contrast in the amplitude of the light waves coming from the object, and those that are transmitted unchanged through the object. This imaging technique may be called amplitude contrast.

5.4.2 Dark field microscopy

In the case of dark field microscopy the background illumination is cut off completely, leaving the field of the image dark. An extra condenser lens is added so that the rays of light fall obliquely on the object. The direct beam then passes straight through and is cut out by means of a diaphragm. Only the light scattered by the object is collected by the objective and sent on to the eyepiece. The light coming from the source itself has its central portion cut off and only an annular ring of light falls on the condenser lens. This technique requires a very powerful source of illumination since only the scattered light is used in the image formation. Nevertheless it is one of the oldest kinds of microscopy and still finds application when viewing very small objects. Bacterial flagella, for example, with diameters of only about 20nm can be followed in motion using a very strong source of light and dark field microscopy. This is also a form of amplitude contrast microscopy.

5.4.3 Phase contrast microscopy

Every wave has a characteristic phase. This is illustrated in Figure 5.7 where wave b has the same phase as wave a while wave c has a different phase. In a microscope if the phase of the rays scattered from the sample can be made to have a different phase from the unscattered rays, then the image of the specimen will be different in light intensity as compared to the background and will be more clearly visible. This is because when two waves (e.g. the scattered rays and unscattered rays) are in phase, they reinforce one another and the total amplitude is larger. Against this, when the two waves are out of phase, they combine together destructively and give a smaller amplitude. This effect is used in a phase contrast microscope to image thin transparent specimens. The basic principle of the method is that the unscattered beam in the direction of the incident light has its phase shifted (or changed) inside the microscope before being allowed to recombine with the scattered light. The phase

shifting is done as follows. The illumination of the specimen is made obliquely using a condenser lens as in the case of dark field microscopy. Instead of a diaphragm cutting off the direct beam, however, a phase plate is placed in the focal plane of the objective lens. The phase plate consists of a transparent disc in which there is an annulus of smaller optical path than the rest of the plate. Since the medium of the plate is dense, light passing through it will be retarded. The direct beam will be less retarded than the scattered light, thus introducing a phase difference. This enables the image to be formed clearly in a bright field. Two effects usually accompany phase contrast microscopy. First, some thick objects may show less contrast than thinner ones and may even reverse contrast so that they appear brighter than the background. Second, every object is surrounded by a diffuse halo of light, and does not represent the real structure. These effects are due to the fact that, apart from the phase plate, the objects themselves introduce a shift in phase. In the case of the thick objects the shift in phase may be such that when the rays recombine, the scattered and the unscattered light combine constructively, thus producing reverse contrast. The halo effects appear because some of the direct beam, instead of passing through the retarding portion of the phase plate, passes through the other portion. This creates the effect of a bright, blurred, low-resolution image of the object appearing superimposed on the sharper actual image. The phase contrast microscope is used only with bright field illumination. However by changing the shape of the phase plate such that, instead of a ring of reduced thickness for the unscattered beam, we have a ring of enhanced thickness, it is possible to have reverse contrast for thin objects instead of positive contrast.

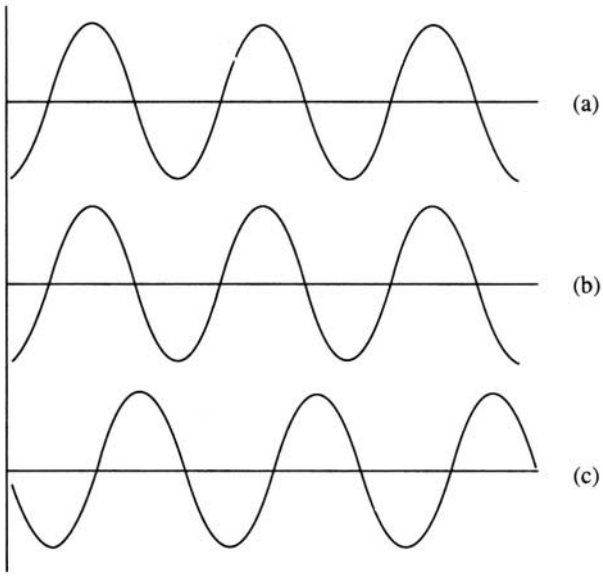


Fig. 5.7 Waves of different phases. (a) and (b) are in phase and (c) is 90° out of phase with the other two

5.4.4 Fluorescence microscopy

Fluorescence is the phenomenon by which certain substances absorb light of a particular wavelength. After a very short interval of time the light is re-emitted with its wavelength altered. If the delay between absorption and emission is less than 10 seconds, the effect is called fluorescence. Longer delays lead to the phenomenon of phosphorescence. Both go under the general name of luminescence. In general the wavelength of the emitted light is longer than the wavelength of the absorbed light.

Fluorescence microscopy uses this property to specifically image materials of interest While performing the experiment, filters are used for both the incident light and the scattered light so that only a narrow band of wavelengths is passed in each case. The incident light filter allows only those wavelengths that will be absorbed to impinge on the specimen. The filter on the scattered light, also known as the barrier filter allows only the emitted wavelengths to pass through. Since the emitted wavelengths are different from the absorbed wavelength, all the background is filtered out and only the image of the sample is allowed through. Many natural materials exhibit fluorescence. The amino acid tryptophan, for example, is an ubiquitous component of proteins and fluoresces in the 250 to 400 run wavelength region. Such an auto–fluorescer, however, is of limited value.

Of much greater value is the use of flourochromes. These are small fluorescent molecules, which may be linked chemically to materials of interest. Fluorescein isothiocyanate (FITC) absorbs blue light and emits green light. Rhodamine compounds absorb green light and emit red light. These compounds are especially useful in immuno–fluorescence microscopy. In this technique fluorescent dyes are linked to antibodies raised against the molecules of interest. The labeled antibodies are then cross-reacted with the sample containing the antigens. The fluorescent image of the sample will then clearly indicate the distribution, and even the size and shape of the antigen. In clinical medicine, fluorescence microscopy is a powerful technique to identify the presence of particular viruses in the sample. In biology, it enables the mapping of the distribution of even fairly small cell components. The antibodies are raised directly against the material extracted from similar cells and purified biochemically. Another method is to attach the fluorochrome to an anti-antibody, i.e. a general purpose antibody which will attach to the antibodies against the particular materials. By this method, the process of attaching the ‘stain’ to the antibody each time is avoided as the non-specific, labeled anti-antibody can be produced commercially and stored.

5.4.5 Polarising microscopy

As electromagnetic radiation, such as light, propagates, its electric field oscillates perpendicular to the direction of propagation. If the electric field is confined to one plane (say, the vertical ‘up-down’ plane) the light is said to be plane polarised. Light can be plane polarised by crystals or by special materials such as Polaroid. When a beam of unpolarised light falls on a sheet of Polaroid, only light polarised in one particular direction passes through. The rest is absorbed. Thus when a beam of polarised light falls on such a sheet, whether the light passes through or not depends on the relative polarisation directions of both the beam and the sheet. A device that polarises the light is called a polariser. A device used to analyse the direction of polarisation of the beam is called an analyser. Both the polariser and the analyser are normally the same material such as a nicol prism or a sheet of Polaroid.

In a polarising microscope, light polarised by the polariser is allowed to fall on the object. Light scattered from the object is then viewed through an analyser. A polarising microscope helps to determine the structures of ordered domains within the sample. This is because such ordered domains possess the property of birefringence. In stretched biological fibres, for example, most of the chemical bonds will be oriented in a particular direction and the electric field of the incident light interacts very strongly if it is polarised in the same direction. In effect the fibres act as a kind of polariser. This property is called birefringence. If the direction of oscillation is along the fibre axis, the light travels slower than when the electric field oscillation is perpendicular to the fibre axis. This is called positive birefringence. For negative birefringence, the long film axis is the fast axis and the shorter axis is the slow axis. If therefore a beam of plane polarised light is incident on a birefingent object the effect is to turn the plane of polarisation by an angle. If now the object is viewed through the analyser placed

perpendicular to the polariser, all the direct light is cut off and only the light from the object will pass through (actually only a component of this light will pass through), and the birefringent object is clearly visible. By rotating the sample and observing the change in the intensity, it is possible to measure very precisely the direction of orientation of the birefringent portions of the sample. Typical applications of the polarising microscope in biology include the study of cell spindle formation and the behaviour of the components of a contracting muscle. Polarisation measurements can also be made in conjunction with fluorescence microscopy and information about the orientation of the marker dyes is obtained.

6

Electron Microscopy

6.1 Introduction

Electrons are negatively charged subatomic particles. Their use in imaging can be explained in terms of their quantum mechanical wave-like property, which allows a beam of electrons to be considered in exactly the same way as a beam of electromagnetic radiation such as light. Electrons possess two additional advantages for use in microscopy. The first of these is that according to quantum physics, the wavelength of the waves associated with an electron varies inversely as its energy. Thus by accelerating the electron between charged plates with a very high applied voltage, the wavelength of the 'radiation' used in the microscope could be made very small. As explained earlier, this implies a concomitant increase in the resolution and hence in the possible magnification. The second advantage, which actually allows the use of electrons for microscopy, is that, unlike electromagnetic radiation of short wavelengths such as X-rays, electrons can be focussed by means of magnetic lenses. This is dependent on the well-known principle in physics that a charged particle in a magnetic field is deflected in a direction perpendicular to its direction of motion. This principle has been used in the construction of electron microscopes.

In this chapter we will first study some elementary electron optics including the principle components of the electron microscope. Next we describe the different types of microscopic techniques, the preparation of the specimen, and some applications. We go on to an introduction to advanced image reconstruction techniques and electron diffraction. We end with a discussion of tunnelling electron microscopy and the atomic force microscope.

6.2 Electron Optics

The wavelength of the electron beam depends on its energy, which is in turn dependent on the voltage used to accelerate the electrons. Modern electron microscopes use accelerating voltages in the range 1000 volts to 1000 kilovolts. The most popular microscopes use about 100 kilovolts. According to the

de Broglie equation (see Chapter 1), this voltage will correspond to a wavelength of about 0.03 Å. This should, in principle, allow the imaging of details on an atomic scale. However several factors come in the way of reaching this high resolution. Vibration, damage to the specimen and contamination are some of the problems encountered in microscope design. The most important factor limiting the resolution is the inability of the magnetic lenses to focus the beam accurately. This is called the aberration of the lens. Aberration is chiefly of two types. The first, called spherical aberration arises from the fact that all the electron beams scattered from the specimen which fall on the lens should be focussed on to the same spot but in practice are not. Figure 6.1 gives an illustration of how the images from a point will be spread out into a series of images. Thus the image of the point will be seen as a disk, called the disk of confusion. The spherical aberration can be reduced by cutting out some of the higher angle beams scattered from the point by means of a stop as illustrated in the figure. But obviously there is a limit to the narrowness of the aperture beyond which diffraction effects will come into play. The second important type of aberration, called chromatic aberration arises from the fact that the focal length of the lens is different for different wavelengths. Since the source cannot produce a uniformly energetic electron beam, therefore the image of the point again spreads out into a disk of confusion, this time caused by the chromatic aberration.

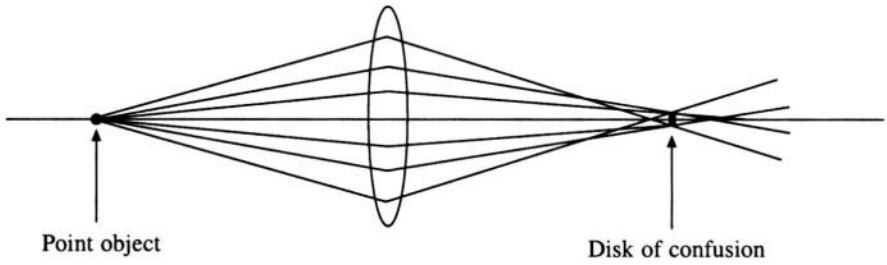


Figure 6.1 Spherical aberration in electron optics, caused by imperfections in the magnetic lens

The final quality of the image depends not only on the resolution but also on the contrast. Contrast is the difference in intensities between the object and the background as a fraction of the background intensity. For visibility, this ratio should be larger than 0.2. The contrast can be enhanced if the object is dense as compared to the background. Usually, the support for the specimen is a carbon film and biological materials, especially, have about the same density. The use of electron microscopy in biology is thus especially problem prone. As will be described later, one can use staining methods to increase the contrast. It is also possible to increase the contrast by increasing the intensity of the incident electron beam. This is because the background intensity varies as the square root of the incident intensity while the image intensity will vary directly as the incident intensity. However increasing the intensity of the incident beam leads to another severe problem in the study of biological materials, viz. radiation damage. This will destroy the sample leading to very poor quality images.

Thus because of the above problems and also because of others not fully discussed here, the actual practical resolution and magnification in biological electron microscopy is below what in principle is possible. Nevertheless, electron microscopy is far superior to the light microscope in its resolution and magnification. Objects such as cells, viruses, DNA strands, ribosomes, etc. can be imaged clearly at resolutions of a few tens of Angstroms, at magnification of 100000 X or more.

6.3 The Transmission Electron Microscope (TEM)

This is one of the most commonly used instruments. Its geometry is shown in Figure 6.2. The source

S of the electron beam may be a tungsten filament, but other materials like lanthanum hexaboride are also used. The beam path, the placement of the lenses, the specimen, the aperture etc, follow the plan of the light microscope as seen in the figure. The entire arrangement, including the specimen, has to be placed in high vacuum to avoid extraneous scattering and absorption of the electrons by air. The lenses are magnetic lenses in the electron microscope and unlike as in the light microscope, the image is not viewed directly through an eyepiece, but is projected on to a fluorescent screen on which the electrons form the image. An extra aperture called the objective aperture is placed in the electron microscope and not normally in the light microscope. This is because the image forming process in the TEM is different from that in light microscopes. In a light microscope, light is transmitted or absorbed by the specimen. This creates a contrast between the different parts of it and hence the image. In the TEMs the situation is quite different and most of the incident beam passes through the specimen unimpeded and very little is absorbed. Figure 6.2 shows that only a small portion is actually scattered in the forward direction. It is this forward-scattered fraction that is used to form the image. Contrast is created by differentiating between electrons scattered into wide angles from those scattered into small angles. The objective aperture cuts off the large angle electrons, and the amplitude of the electron beam from different portions of the specimen is different leading to amplitude contrast. Most biological materials however do not have enough contrast to be viewed unless they are stained. Dark field imaging is also possible and is achieved by cutting out the unscattered beam.

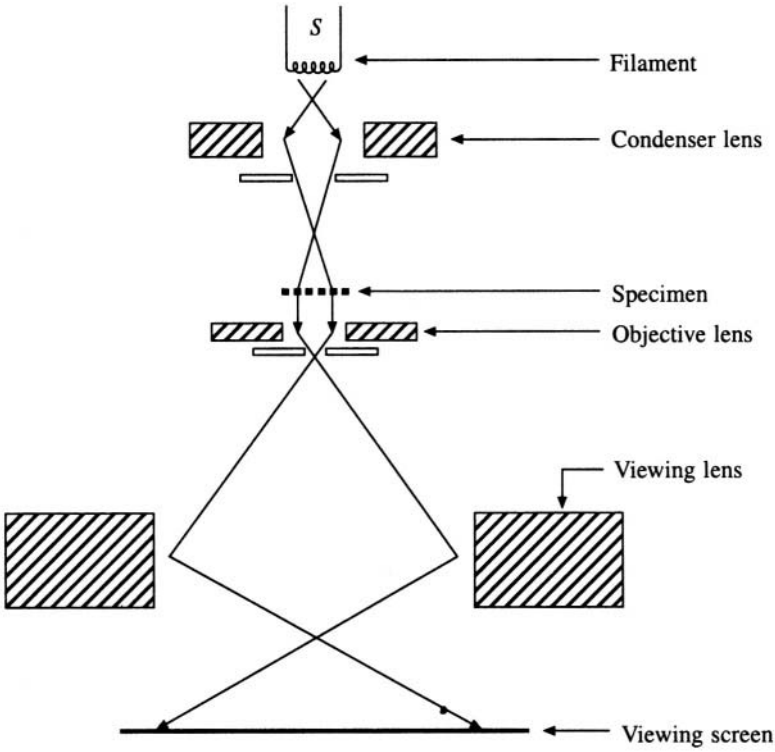


Figure 6.2 Schematic diagram of a transmission electron microscope

6.4 The Scanning Electron Microscope (SEM)

A SEM is very useful in obtaining images of the surface of thick specimens. In the scanning-

transmission mode (STEM), the SEM can also be used to study thin specimens and in this respect has some advantages over the conventional TEM. Figure 6.3 is a schematic illustration of the construction of the SEM. The most obvious difference as compared to a TEM is that the electron beam, after passing through the condenser lens is deflected in a raster pattern over the specimen stage, similar to the pattern in a television picture tube. The objective lens is split into two parts. One part is placed between the condenser lens and the specimen and can in fact be regarded as an additional condenser lens, which focuses the electron beam onto a small spot on the specimen. Image signals can be collected by the detector *D,* which is placed on the same side of the specimen as the source and the raster coils. This detector is used in the surface-scanning mode and collects the secondary electrons knocked out of the specimen by the primary beam, as well as the back scattered electrons reflected from the surface of the sample. Image signals can also be collected in the scanning transmission mode by the detector *D* after passing through the second half of the objective. These are the forward-scattered electrons.

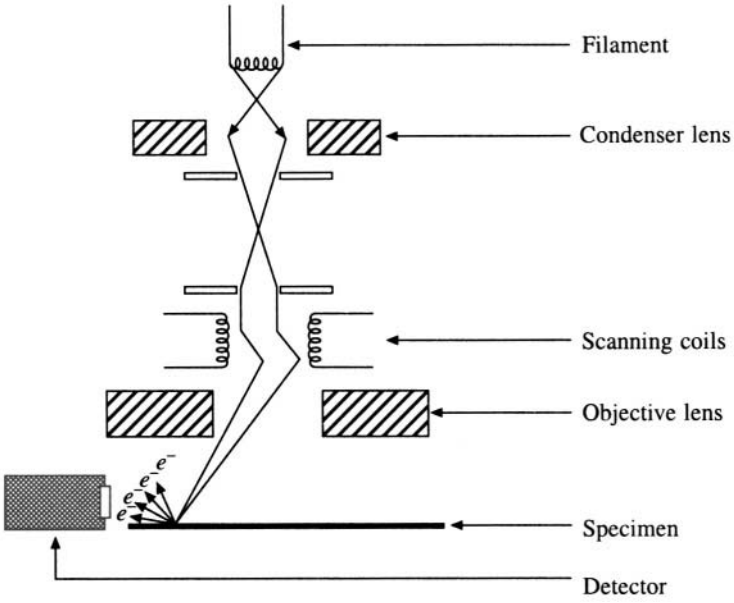


Figure 6.3 A schematic representation of a scanning electron microscope

Images obtained in the STEM mode are very similar to those obtained in the fixed beam TEM. The surface SEM image is quite different. In particular, the contrast in the surface imaging arises partly from the different scattering powers of the different atoms in the specimen, and partly from variations in the topography. In both the STEM mode as well as the surface SEM modes the images are generated in the following way: As the electron beam scans the specimen, the scattered electrons from each position is measured by the detector, one after the other in time. The measured intensities are displayed on the imaging screen one after the other in space, in the same type of raster scan as the electron beam. Thus if the scattering is high at particular point during the scan, then the corresponding point on the viewing screen is bright. If the scattering is low the corresponding point is dark. This creates an image of the object. The resolution of the image depends on the size of the electron spot used to scan the specimen and this may be as small as 50 Å. Both surface SEM and STEM have changed the use of electron microscopy quite dramatically and have made quantitative studies possible.

The scanning principle has also been used more recently in the construction of the atomic force microscope and the scanning tunnelling microscope. These will be discussed later.

6.5 Preparation of the Specimen for Electron Microscopy

Preparation of the specimen is one of the most important steps in the process of imaging by electron microscopy and often takes longer and requires greater skill than the actual observation. This is especially true in the case of biological samples. This is because biomolecules and assemblies are prone to disintegration in the high vacuum condition under which the imaging is done. Secondly without staining with heavy atoms, the lighter atoms such as carbon, hydrogen, nitrogen and oxygen, which are the principle components of biological samples, do not offer sufficient contrast against the support, and cannot therefore be observed. Finally, unless protected, the intense beam of electrons can easily destroy the sample.

The specimen preparation involves the following steps: (a) Preservation: The specimen has to be preserved in the relevant configuration. Care should be taken to avoid artefacts being introduced into the specimen during the preparation. (b) Staining: In order to increase the contrast between the sample and the background, the specimen has to be stained with a heavy, electron-rich metal. (c) The specimen has to be dried, i.e. water has to be removed in order that this does not interfere with the imaging. (d) The specimen has to be made into a form suitable for observation by replication and/or microtomy.

(a) *Preservation*: The two main methods of preserving the sample are by freezing and by chemical fixation. Freezing the sample has to be done with care to avoid formation of ice crystals that may radically alter the structures of interest. Rapid freezing is one way of reducing ice crystal formation and a cooling rate of about 5000 degrees per second is sufficient to freeze whatever water there may be in the sample into a glass like structure which does not affect the sample. Though the sample is usually cooled down to only liquid nitrogen temperatures, such cooling rates may obtained if the whole process is accomplished in a few milliseconds. Once the specimen is frozen and is retained at the liquid nitrogen temperature or below, the structure, morphology and constitution do not alter even for months. The sample can also be preserved by chemical fixation. Chemicals, most commonly osmium tetraoxide and gluteraldehyde are used to preserve the structure and shape of the specimen. Both chemicals work by creating covalent bonds between different molecules, especially proteins or lipids. Osmium tetraoxide has the additional advantage of providing a heavy metal stain, doing away with the necessity of applying the stain separately. However, unlike gluteraldehyde, it does not act on a wide range of tissues. The great advantage of chemical fixation, as compared to freezing is that the procedure is simpler. Moreover, chemicals can be used to preserve much thicker samples since the penetration is greater. Handling and storage of chemically fixed samples is also easier.

(b) *Staining*: As mentioned earlier, both the biological samples studied and the carbon films on which they sit consist of light atoms and without contrast enhancement, the samples will be invisible with respect to the background. Thus staining the sample must be resorted to, in spite of the fact that this will result in a substantial loss of resolution. The technique of positive stain adds a specific heavy atom (uranium to DNA samples, or osmium to lipids) by chemically complexing them. Another technique of positive staining is called shadowing. Here heavy atoms, such as tungsten are evaporated from a filament onto the sample. If the sample is bombarded at an angle then walls of the metal atoms pile up on one side of vertical features in the sample, while no metal atoms are deposited in the shadow of the wall. This allows the easy visualisation of the contours of the raised features in the sample. In the technique of negative stain, the heavy metal stain fills the area where the molecules are not present. Thus the outline of the molecules or molecular complexes will be visible.

(c) *Dehydration*: For wet specimens the dehydration has to be carried out gradually in order that the sudden loss of water does not distort the structure. In the case of frozen specimens such sudden change in the water content is less likely, and dehydration is carried out by maintaining the sample in an atmosphere of very low pressure. A common method of dehydrating wet samples is to immerse it in alcohol solutions, slowly increasing the concentration of alcohol, such that the water is ultimately replaced completely by alcohol. In some cases the heavy metal stain can act also as the dehydrating agent. When a thin layer of the staining solution such as potassium phosphotungstate or ammonium molybdate or uranyl acetate is added to a thin specimen exposed to air, the solution takes on an amorphous, glassy shape and protects the morphology of the specimen during dehydration.

(d) *Replication, microtomy etc*: This step in the specimen preparation process renders it suitable for viewing in the microscope. Biological macromolecules such as DNA etc. can be deposited directly onto a very small (~1 mm diameter) fine grid made of copper which is then treated with the staining solution and inserted into the microscope for viewing. Replication is a method used to prepare samples when only the surface features are of interest A thin layer of metals such as gold, tungsten, palladium or platinum, or a thin plastic film is cast over the sample and allowed to set. The specimen can then be discarded, as a replica of it is now preserved in the metal or plastic film. If the interior of the specimen is of interest, sectioning becomes necessary. In this case the specimen is embedded in a resin such as epoxy resins ('araldite') or polyester resins. It is then sectioned using a microtome. A microtome is a device that uses glass or diamond knives to produce very thin sections of the embedded samples. The thickness of the sections can be as small as 100 nanometers. These sections are then placed on the copper grid for viewing. The heavy metals that are used in the staining process also act to conduct away the excess negative charge that accumulates on a non-conducting sample. For very thin samples viewed on the copper grid, the grid will act to ground the excess charge. In cases where a thick non-conducting sample is used without staining, e.g. for surface SEM studies, it should be coated with a thin metal film solely to act as the conductor.

6.6 Image Reconstruction

Very often the electron micrographs of biological samples such as viruses or ribosomes or other large molecular assemblies consist of several repeated indistinct images of the object. The image can be made more distinct by reconstruction procedures. There are chiefly two such procedures. The first involves simply superposing all the images, such that the 'true' parts of the images, or the signal, gets reinforced, while the background 'noise' will be cancelled. The selection of the identical parts of the image for superposition can be accomplished visually. A more rigorous method is to calculate a superposition function in two dimensions. The superposition function or the cross correlation function will have a maximum value when one image is rotated and translated to superpose on another image optimally. Such optimisation can be accomplished on the computer if the images are digitised and fed to it. A more accurate way to accomplishing a clear image from several indistinct images of the same object is possible if the images are arranged in a regular repeating two-dimensional pattern. If the image is used as a kind of imperfect diffraction grating and is placed in the path of a source of light, a diffraction pattern will be formed. The diffraction pattern can then be filtered to allow through only those parts, which correspond to the periodic lattice. The filtered diffraction pattern is in turn used as the grating and placed in the path of a light beam. The image formed will correspond to a much clearer version of the original object. Mathematically, the procedure is equivalent to performing a Fourier transform, cleaning up the coefficients by filtering, and then performing a reverse Fourier transform. The method of image reconstruction has been widely use to study crystalline materials. Virus structures such as the stem of the *T* phage, membranes and two-dimensional arrays of ribosomes

have all been studied by this method. The method can be extended to three dimensions by taking several two-dimensional images obtained by tilting the specimen at different angles. In fact a single micrograph of a suspension of particles is bound to contain images of the particle in various three-dimensional orientations. Image reconstruction techniques can be fruitfully used to put together these images to form a 3-*D* picture of the particle.

6.7 Electron Diffraction

The electron microscope can be also be used with a few modifications, to obtain electron diffraction patterns from two-dimensional crystals. Instead of analysing these patterns through geometrical optics, it is possible to use Fourier theory to mathematically transform the diffraction pattern into a three-dimensional image of the particle. Since the electron wavelength is much smaller (~0.03 Å) than that of X-rays, it may be thought mat a much higher resolution will be obtained. However several obstacles come in the way of attaining such high resolution. Since electrons are easily absorbed, only very thin samples can be Used. Electron diffraction is ideal for use in those cases where three-dimensional crystals do not grow but two-dimensional arrays form easily, such as to study membrane proteins. From such a thin 'crystal', almost the entire diffraction pattern will be in a single plane and most of the three-dimensional pattern required to perform a complete Fourier transform is not available. Another problem, also present in the case of X-ray diffraction, is the loss of phase information and the consequent inability to perform the Fourier analysis. One way in which this problem is got over is by using the electron micrograph to obtain some information about the approximate phases and using this information in combination with the measured electron diffraction intensities. In spite of these difficulties, electron diffraction is a popular technique to obtain low-resolution images of large particles that do not easily form crystals suitable for X-ray diffraction.

The experimental techniques of preparing the sample and obtaining the diffraction pattern are not very different from those for obtaining electron micrographs. Since the electron beam can be focussed on very small areas, the ordering in the sample needs to extend only up to a few microns (or less) to obtain the pattern. The intensities are then extracted from the electron diffraction photograph by densitometry. The intensities and the calculated or estimated phases are combined to perform a Fourier transform to obtain the electron density images of the particles. It is possible to obtain the diffraction patterns from tilted crystalline samples and calculate three-dimensional images. These will improve the quality of the final results.

6.8 The Tunnelling Electron Microscope

In the last decade, the tunnelling electron microscope has become an integral part of any laboratory interested in the study of surfaces. Its use in biological imaging is somewhat less widespread, though in recent years it has been used to produce images of DNA molecules that have enabled precise measurement of the pitch of the double helix, as well as of proteins and macromolecular assemblies.

The operation of the tunnelling electron microscope is based on the well-known quantum mechanical principle of electron tunnelling. The equations of quantum mechanics assign a finite non-zero probability for an electron to go from one region of low potential energy, to another such region, even if it does not possess sufficient energy to overcome the barrier between the two regions. This phenomenon may be equivalently visualised as a football going from one well into which it has been dropped into another one nearby, without coming to the surface or making a hole between the wells. In the tunnelling electron microscope, this phenomenon is used in the following way. A very, very sharp needle is brought close to the surface to be imaged. The distance is of the order of a few Angstroms. At such a distance electrons from the surface will tunnel across the gap and set up a measurable

tunnelling current in the needle. This current is a precise indicator of the distance between the tip and the surface. The needle can now be moved (in a raster pattern) across the surface, at each point its height being adjusted to maintain the tunnelling current at a constant value. Thus a record of the vertical portion of the needle at each position becomes equivalent to a record of the surface topography and can be converted into an easily observable picture.

Figure 6.4 is a schematic representation of a scanning tunnelling electron microscope. The tip of the needle can be made so fine that at the very end there is only one single atom. Also, modern electrical measurement technology allows even nano-amperes of tunnelling current to be measured. These technological developments have been instrumental in making possible the visualisation of individual atoms on the surface of materials.

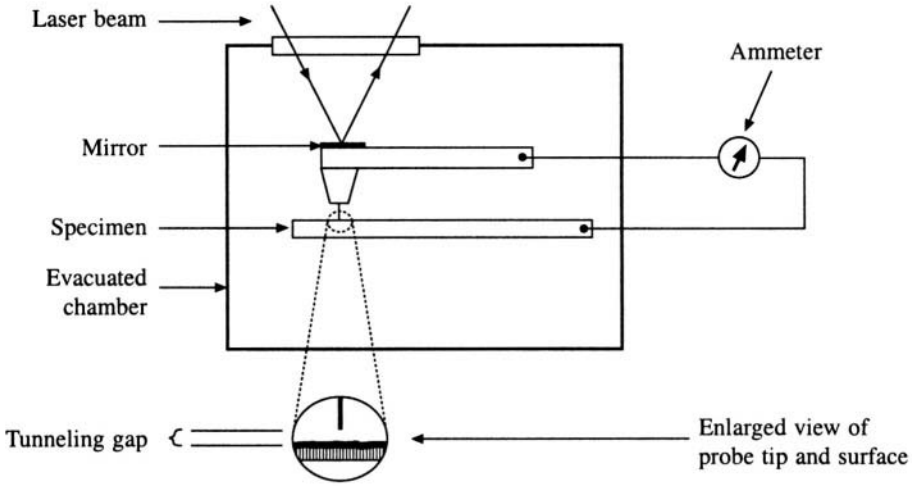


Figure 6.4 A schematic diagram of a scanning tunneling electron microscope

The use of this microscope in the study of biological samples is still not widespread for several reasons. The chief among these is that non-conducting or semi-conducting materials have to be stained with heavy metals before they can be imaged, leading to a loss of resolution. The preparation of the tip and its positioning above the surface also has to be more precise and controlled in the case of biological materials.

6.9 Atomic Force Microscope

The tunnelling microscope has been modified in the following way. Instead of keeping the needle tip at a small distance from the sample the needle is pushed right against the surface. The force present in the tip is kept constant and the tip is scanned across the surface like a phonograph needle running in the groove of a record. A record of the tip's vertical motion is made by reflecting a laser beam from a mirror fixed to the top of the needle and using an interferometer to accurately measure the distance travelled by the laser beam, and hence the position of the tip. This record can be converted and displayed as an image of the surface. Since the atomic force microscope does not depend on a current, it can be used to visualise the surfaces of conductors as well as non-conducting materials. However viewing biological samples such as cells or proteins is still a problem, as the tip tends to distort them.

Other microscopes have been built which are based roughly on the same principle of a tip scanning

the surface. In each different type of microscope a different force is measured. Thus we have friction force microscopes, magnetic force microscopes, electrostatic force microscopes, and scanning thermal microscopes. The last may be described as the world's smallest thermometer. A thermocouple attached to the needle tip measures the temperature at each position. Since heat is essentially motion at infrared wavelengths, it may soon be possible to measure the infrared spectrum of a single molecule!

7

X-ray Crystallography

7.1 Introduction

In June 1913 Laurence Bragg determined the structure of NaCl using X-ray analysis. Since then, in the intervening 90 years, continuous development of theory, experimental practice and technology has made possible the determination of the three dimensional structures of nearly 250,000 organic and biological molecules. This chapter discusses the powerful technique of X-ray crystallography in detail. Crystallography is the study of crystals. Since biological materials are only rarely crystalline in the natural state, one may wonder what biophysics has to do with crystallography. The answer lies in the fact that, except for NMR, X-ray crystallography is the only available technique of obtaining a detailed, accurate and unbiased picture of a molecule.

The method of X-ray crystallography (Figure 7.1) can be described very briefly as follows: Crystallise the molecule, subject it to irradiation by a beam of X-rays and make a record of the 3-D diffraction pattern. Analysis of this data through computer calculations results in a model of the molecule, which can then be refined against the observed X-ray data. Finally one obtains the three-dimensional coordinates, which describe the position in space of each atom in the molecule. Such a refined model can be displayed and manipulated either as plastic or wooden models, or on the computer graphics terminal. In the following portions of this chapter we will look into each of these aspects in detail.

7.2 Crystals and Symmetries

A crystal is a regular, repeating three-dimensional array of atoms or molecules or groups of molecules. If one imagines a point as representing each repeating unit, the array of points which represents the crystal is called a lattice (Figure 7.2). There are 14 different symmetrical patterns in which the points can be arranged in space and in addition be consistent with the requirement that the patterns be repeated in all three dimensions regularly to form the crystal. In other words, whatever other symmetry

the lattice may possess it has to be consistent with the translational symmetry that defines a crystal. These 14 patterns are called Bravais lattices.

The Bravais lattices exhibit mirror symmetry and rotational symmetries. A lattice possesses mirror symmetry, if there exists in the crystal a plane across which the lattice can be reflected without in any way changing the pattern of arrangement of the lattice points. Similarly, an n-fold axis of rotational symmetry exists if the lattice can be rotated about the axis by 360/n° without any change in the pattern. Consistent with translational symmetry, only one-fold, two-fold, three-fold, four-fold and six-fold axes of rotational symmetry are possible.

Each Bravais lattice may be characterised by three axes, which are defined by joining a point to three neighbouring points that are not in the same plane. A unit cell is formed by completing the box and is described by the length of the three sides *a, b* and *c* and the angles between them, α, β and γ (Figure 7.3).

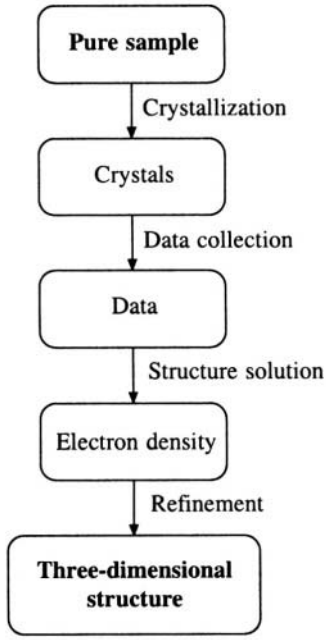


Fig. 7.1 Flowchart of the X-ray crystallographic method to obtain the three-dimensional structure of molecules

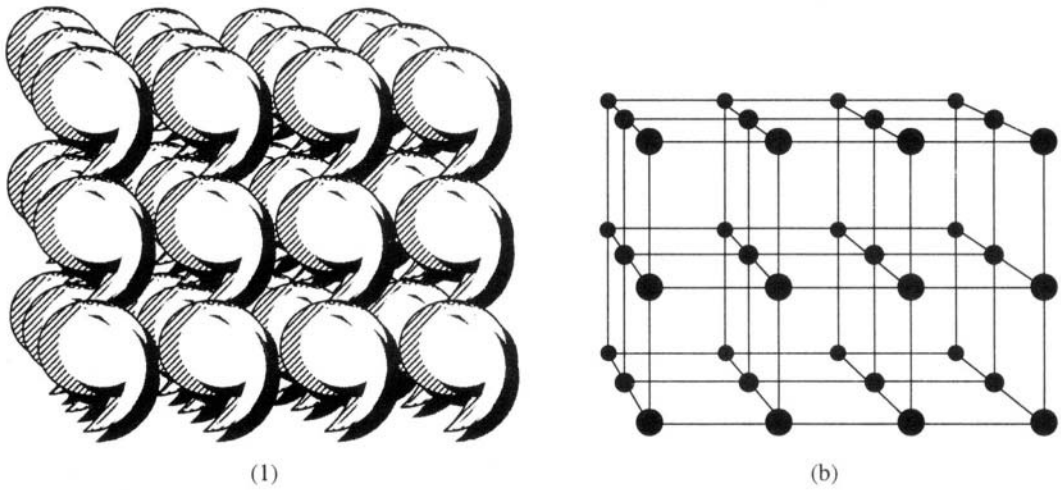


Fig. 7.2 (a) A regular periodic arrangement of three-dimensional objects. (b) The underlying lattice

7.3 Crystal Systems

A triclinic unit cell has three unequal axes and angles (Figure 7.4a).

A monoclinic unit cell (Figure 7.4b) has three unequal axes, two of the angles as right angles, and no limitation on the third one. This condition can be obeyed both in the case where the lattice points

are at the corners of the unit cell, as well as when there is an extra lattice point on one of the faces. This is conventionally taken to be the ab plane or the 'C' face. There are thus two possible Bravais lattices in this system.

An orthorhombic unit cell (Figure 7.4c) has three unequal axes, and all three angles are right angles. There are four possible Bravais lattices; with the lattice points at the corners; with an extra lattice points on the ab plane; with an extra lattice point at the body centre of the unit cell; and finally with extra lattice points on all six faces.

A rhombohedral system (Figure 7.4d) has three equal sides and three equal angles.

A tetragonal system (Figure 7.4e) has two equal sides and all angles as right angles. There are two possible Bravais lattices.

A hexagonal unit cell (Figure 7.4f) has two equal axes, and one angle as 120°, the other two as right angles.

A cubic system (Figure 7.4g) has equal sides and right angles. There are three Bravais lattices.

The Bravais lattices represent a complete set. No others are possible in a crystal. For example, a tetragonal lattice could have points at the centres of all the faces, like the face-centred cubic lattice in Figure 7.4g. However, this is not another tetragonal lattice, for a rotation of 45° around the c axis will give back the body-centred lattice of Figure 7.4e.

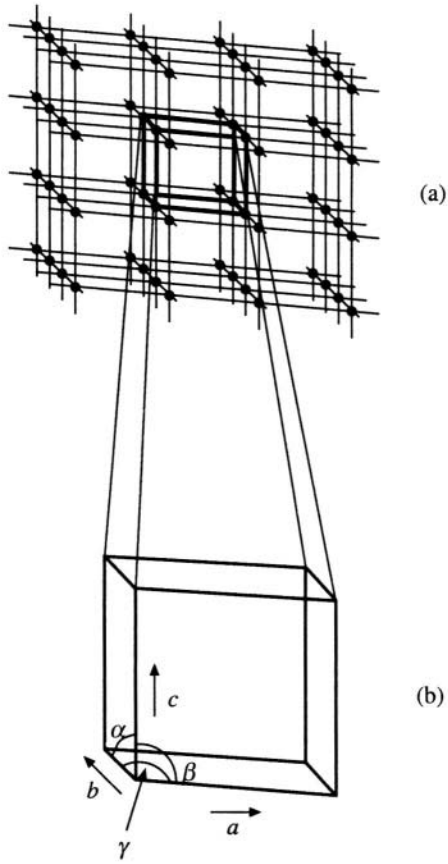


Fig. 7.3 (a) The lattice (b) The unit cell

7.4 Point Groups and Space Groups

Crystals can be classified in terms of the symmetry between their faces. In this case we consider the crystal as a whole and ignore for the present the lattice structure underlying it. Each crystal can be identified by the group of symmetry operations that pertain to it. The symmetry operations can be n-fold rotation, and mirror planes. In addition we may have centres of inversion—each point in space is matched by a similar one at an equal distance on the opposite side of the centre. Rotatory inversion is another symmetry operation that the crystal as a whole may possess—the operation consisting of rotation about an axis combined with inversion about a centre.

Each group of symmetry operations that identify a crystal is called the symmetry point group of that crystal. Since the point group of the crystal is also the point group of the underlying lattice, only symmetry operations consistent with lattice translations are allowed. With this restriction, a total of 32 point groups is possible for crystals and is divided among the seven crystal systems. For example, the monoclinic system has always one two–fold axis, and therefore the point groups 2, (i.e. one two-fold axis alone), m (i.e. one mirror plane alone) and 2/m (i.e. a two-fold axis and a mirror plane perpendicular to it) all fall in the monoclinic system. The point group 222 (three two fold axis

perpendicular to each other) requires three perpendicular axes and hence will fall in the orthorhombic system (Figure 7.5). Other point groups are similarly assigned to other systems.

In addition to the lattice, when we take into consideration the structural components of crystals, i.e. the atoms or molecules, additional symmetry elements, combining translation with rotation or reflection come into play. One of these is the screw axis which corresponds to a rotation and a translation, and is designated $\mathbf{n}_m$ which implies an n-fold rotation followed by a translation of *m*/*n* of the unit cell axis parallel to the rotation axis. Thus, for example, a $\mathbf{3}_1$ screw about c implies a 3-fold rotation about the c axis and a translation of 1/3 times the length of c along the c axis (Figure 7.6). Similarly a glide plane arises when a mirror reflection is followed by a translation of 1/2 along a unit cell axis (*a, b,* or *c* glides) or a unit cell diagonal (*n* glide).

The three dimensional assembly of planes, axes and centres of symmetry is called a space group.

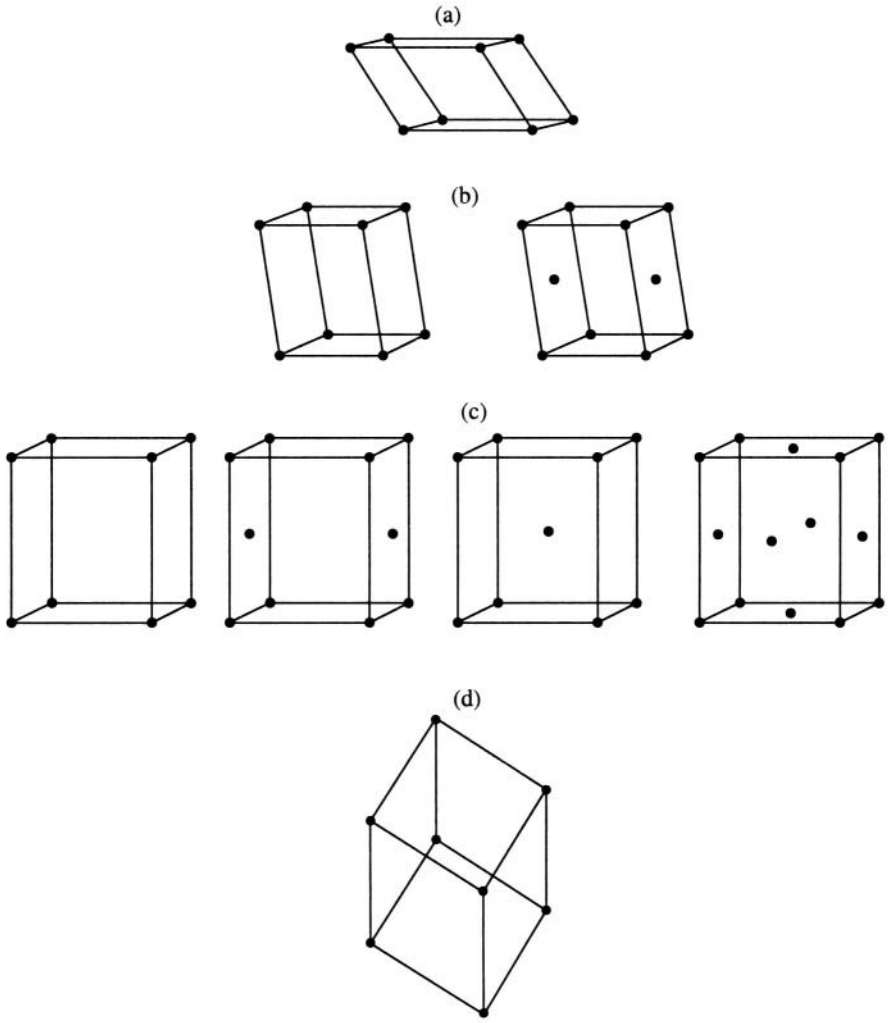


Fig. 7.4 (*Contd*)

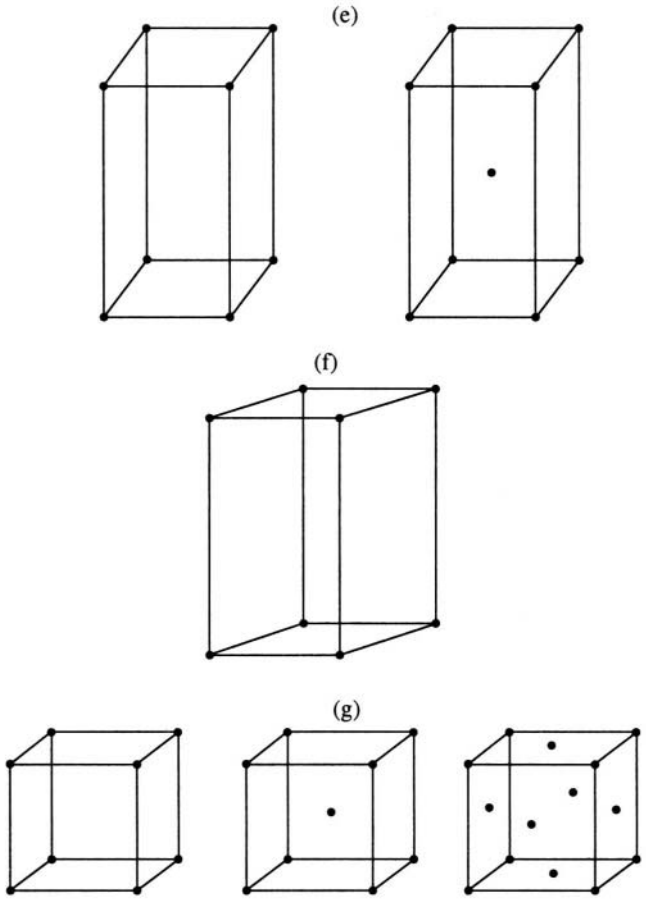


Fig. 7.4 The seven crystal systems and the fourteen Bravais lattices. (a) Triclinic. (b) Monoclinic. (c) Orthorhombic. (d) Rhombohedral. (e) Tetragonal. (f) Hexagonal. (g) Cubic

The symmetry operation of a space group, together with the unit cell translations, will transform a single object into a crystal. Just as there are only seven distinct crystal systems, fourteen Bravais lattices and 32 point groups, there are only 230 space groups, or 230 distinct ways in which an object, such as a molecule, can be arranged in three dimensional space to form a crystal. A description of the 230 space groups is beyond the scope of this book. The International Tables for Crystallography[7] contain all the details. The symmetry of the crystal, i.e. its space group, can be deduced from its X-ray diffraction pattern and is an essential first step in the X-ray analysis. Before we can subject a crystal to X-ray analysis, however, we need to grow it.

7.5 Growth of Crystals of Biological Molecules

The chief difference between crystals of biological substances and those of other important materials such as silicon is that in the former the repeating units are invariably molecules or groups of molecules,

[7] *The International Tables for Crystallography, Vol A,* T. Hahn, (ed.), 1987. D. Reidel Publishing Company, Dordrecht, Holland.

while in the latter the repeating units are atoms or groups of atoms. This implies that the forces of bonding between one repeating unit and another are much weaker in the former, being van der Waals forces or hydrogen bonds. In the latter, on the other hand, the crystals are built up from strong covalent or ionic bonds. Crystals of biological materials are thus usually soft and brittle as compared to inorganic substances.

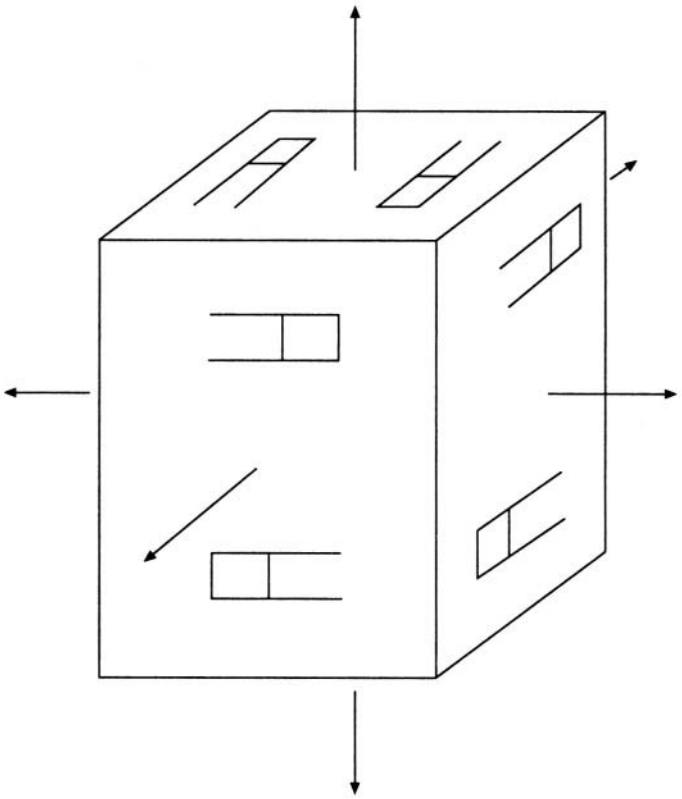


Fig. 7.5 An illustration of the 222 point group symmetry

Another difference is the size. Biological and organic molecules usually form very small crystals (approximately $1\ mm^3$) as compared to inorganic materials (about $1\ cm^3$). Bio-crystals are usually grown from solution and may have a large solvent content. This is especially true of crystals of macromolecules such as proteins and nucleic acids where the solvent can take up as much as 80% of the crystal. The growth of bio-crystals is a multi-parameter process. Table 7.1 gives a list of factors that could affect crystallisation.

To grow crystals of any compound from solution, the molecules have to be brought to a supersaturated state. When the solution returns to equilibrium, the substance will precipitate out, hopefully as crystals, and not as an amorphous precipitate. Supersaturation can be achieved by slow evaporation. Also by varying one of the parameters listed in Table 7.1 one may change the solubility and hence the saturation level of the molecule.

In order to obtain good, defect free crystals of a size suitable for X-ray analysis, the equilibration needs to be very carefully controlled, and techniques have been developed for this.

Slow evaporation (Figure 7.7) is conceptually the simplest of techniques. It is chiefly used to crystallise small organic molecules. The solution is placed in a small glass beaker and sealed with an

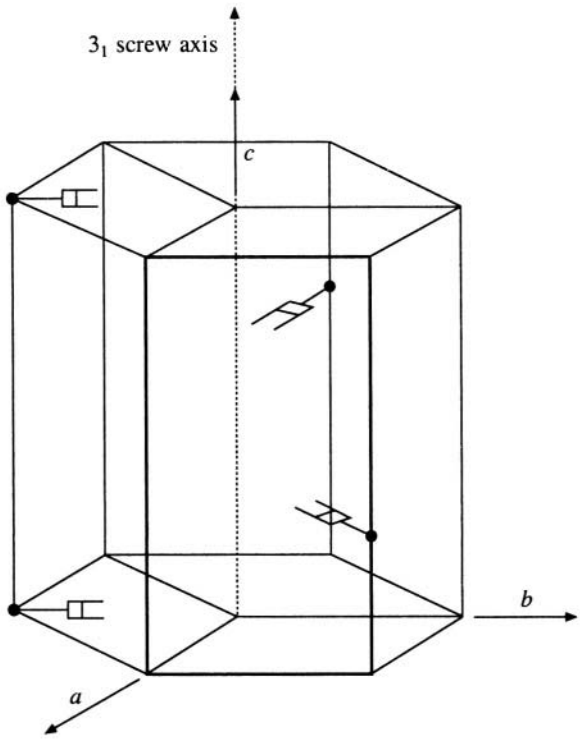


Fig. 7.6 The 3_1 screw symmetry axis in the trigonal system referred to hexagonal axes

airtight cover, except for one or two small apertures through which slow evaporation of the solvent takes place, bringing the solute to supersaturation and hence to crystallisation.

Table 7.1 Factors affecting crystallisation of organic molecules and biological macromolecules

— purity of the sample.
— concentration of the sample
— temperature, pH of the solution
— time, rate of equilibration
— ionic strength
— convection currents
— volume of the crystallisation solution
— pressure, vibrations
— additives, their character and concentration
— solvents (especially for small organic molecules)

For biological macromolecules, especially, dialysis can be used to obtain crystals. The solution of the molecule is separated from a large volume of the solvent by a semi-permeable membrane which gives small molecules (ions, additives, buffer, etc.) the freedom to circulate but prevents the movement of the macromolecule. The macromolecule solution is gradually built up to supersaturation by changing the concentrations of the small molecules in the large volume of the solvent. This has the effect of gradually changing the pH or the ionic strength or any similar parameter.

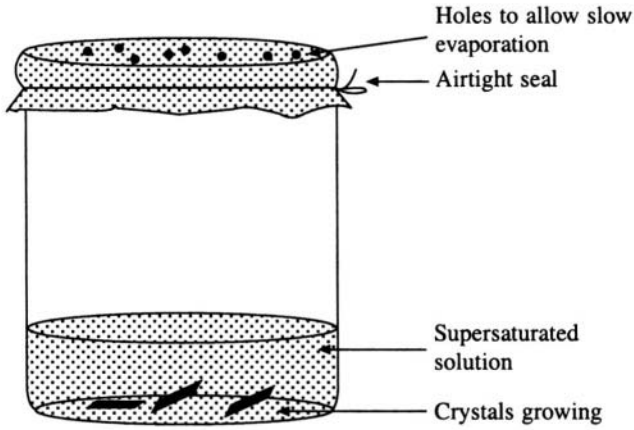


Fig. 7.7 Slow evaporation technique to grow crystals from a supersaturated solution

Vapour diffusion is another commonly used technique (Figure 7.8). A droplet of solution containing the molecule is equilibrated against a reservoir containing the precipitant, i.e. the crystallising agent, at a higher concentration than in the drop.

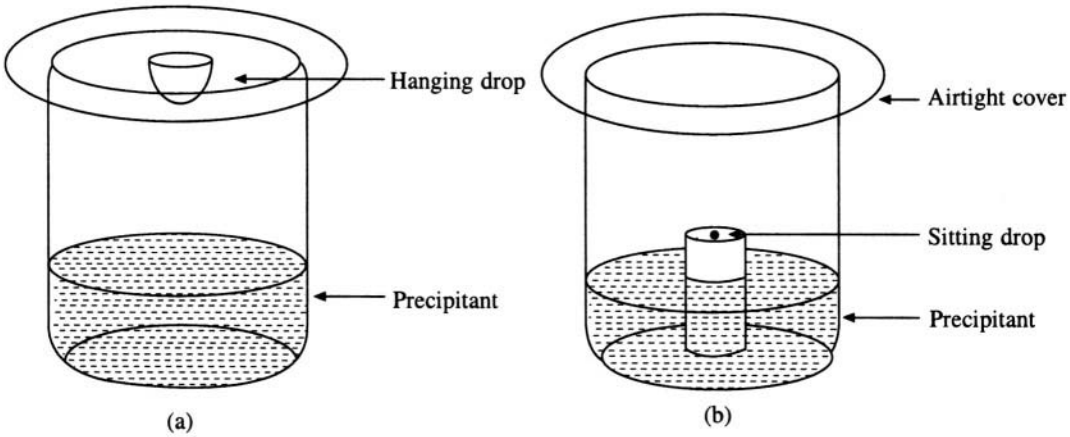


Fig. 7.8 Vapour diffusion method. (a) Hanging drop. (b) Sitting drop

In batch methods, the molecule to be crystallised is mixed with the crystallising agent at a high concentration so that supersaturation is immediately reached. Though this may lead to a large number of tiny crystals, it frequently gives rise to good crystals.

Interface diffusion (Figure 7.9) is used for small molecules as well as large biological macromolecules. The precipitant solution is gently layered over the solution of the molecule. Crystals usually grow at the interface.

Sometimes it is necessary to introduce seeds into the solution to induce the crystals to grow. These seeds are usually obtained from previous less successful crystallisation attempts, and may be extremely tiny crystallites. Table 7.2 gives some of the solvents, additives and precipitants used in growing crystals of organic and biological molecules.

7.6 X-ray Diffraction

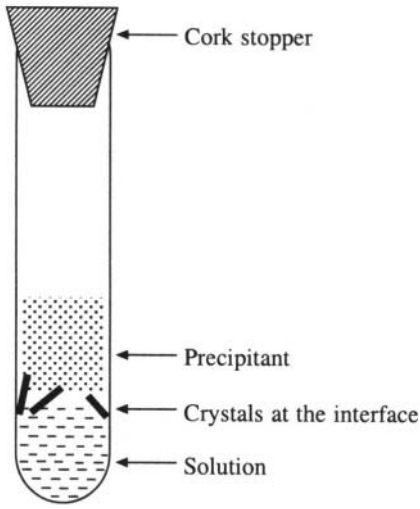


Fig. 7.9 The interface diffusion method of crystal growth

Diffraction is a well-known phenomenon that occurs when two or more waves are combined. If the waves combine in such a way that the peaks of one wave fall on the peaks of the other, the total effect is to enhance the intensity. On the other hand if the combination is such that the peak of one wave falls on the trough of the other, the effect is one of cancelling each other out and the intensity of the combination is zero. If an incoming wave is split into two by slits and the two parts of the wave are allowed to interfere with each other, then a pattern is produced on a screen placed on the other side of the slits. Whether the two beams interfere constructively or destructively at a particular point on the screen, is decided by the difference in the length of the paths that each of the waves take from the slit to the point on the screen. If the difference corresponds to a whole number of wavelengths, the combination of the two waves is perfectly constructive and the intensity is greatly enhanced. A regularly spaced series of slits has the same type of effect and is called a diffraction grating (Figure 7.10). It is possible to have two and three-dimensional diffraction gratings. A crystal lattice is nothing but a three-dimensional grating (Figure 7.2(b)). However, it cannot diffract visible light. This is because for the diffraction effect to be noticeable, the spacing between the slits in the grating should be of the same order of magnitude as the wavelength of the incident radiation. In a crystal the distance between the lattice points corresponds to the dimensions of atoms and molecules i.e., a few Angstroms. X-rays are radiation with wavelengths of this size, and hence diffraction from crystals is observed with X-rays. The beginnings of a theory of X-ray diffraction from crystals was developed by Max von Laue. A few years later, W.L. Bragg came up with the idea of treating diffraction from a crystal as reflection from the lattice planes and put the theory on a firm footing. A crystal can be characterised as having lattice planes, each of which is identified by a set of integers *h, k* and *l* (Figure 7.11). These numbers are known as Miller indices. Miller indices help in identifying the plane from which reflection takes place. The interference of the beams of X-rays reflected from successive parallel planes corresponding to a particular triad of *hkl*

Table 7.2 Compounds that are required to crystallise biomolecules

Molecule	Solvents	Additives	Precipitants
Small organic Molecules	Organic solvents (e.g. CCl_4, ethyl acetate etc.), alcohols, water	Metal ions, salts	Alcohols, organic solvents
Proteins	Aqueous solutions	Co-factors, sugars, metal ions, ion complexes, polyamines, salts	Ammonium sulphate, polyethylene glycol (PEG) Methylpentanediol (MPD)
Nucleic acids	Aqueous solution	Polyamines, metal ions, salts	MPD, PEG, Isopropanol, other alcohols

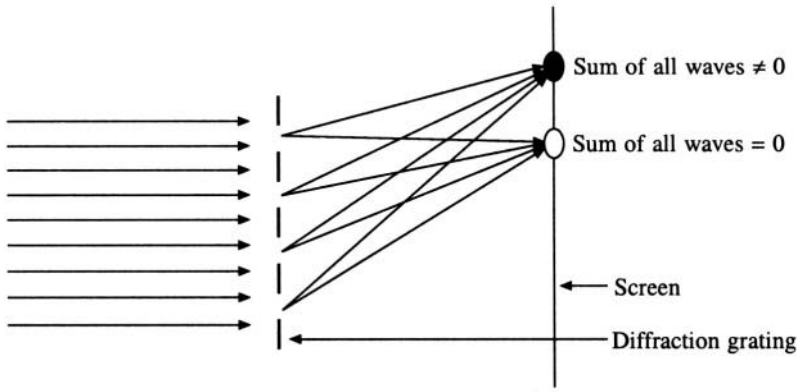


Fig. 7.10 A schematic sketch illustrating diffraction from a grating

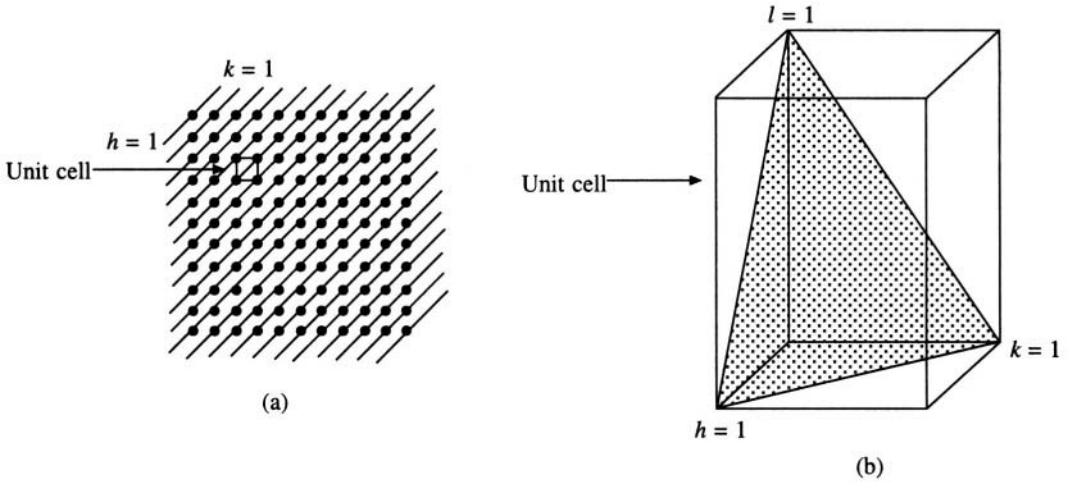


Fig. 7.11 Lattice planes and Miller indices. (a) A two-dimensional lattice showing the set of 'planes' (lines) with the Miller indices $h = 1$, $k = 1$ or (1, 1). (b) A three-dimensional unit cell showing the plane (1, 1, 1)

values, leads to diffraction effects. For the interference to be constructive, as in the case of the one-dimensional diffraction grating, the difference in path lengths of the reflected beams should be equal to an integral number of wavelengths of the incident beam. A simple geometrical construction (Figure 7.12) leads to the famous Bragg reflection condition or Bragg's law

$$2\,d\,\sin\theta = n\lambda.$$

When a crystal is irradiated by an X-ray beam, Bragg's law is used to predict the angles θ at which the diffracted intensities may be expected. This requires knowledge of the inter-planar spacing d and the wavelength λ of the X-rays. Conversely, if the angles of occurrence (and relative intensities) of the diffracted spots are measured in an experiment, the corresponding intes-planar spacing in the crystal may be calculated and the lattice structure can be determined. Therefore, the next step, after growing the crystal is to place it in a beam of X-rays and measure the angles and intensities of the diffracted spots.

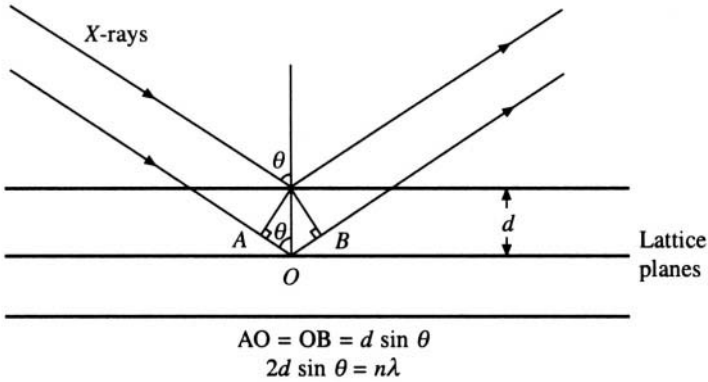


Fig. 7.12 Bragg's Law. For constructive interference to occur, the difference in path lengths, given by *AO + OB* should be equal to an integral number of wavelength or $n\lambda$.

7.7 X-ray Data Collection

The process of collecting a complete set of X-ray diffraction data from a crystal consists of measuring the intensity of the diffracted beam from each set of planes. The experimental set up (Figure 7.13) consists of an X-ray source, a goniometer (Figure 7.14) on which to mount the crystal, and a detector to measure the intensity of the reflected X-ray beam.

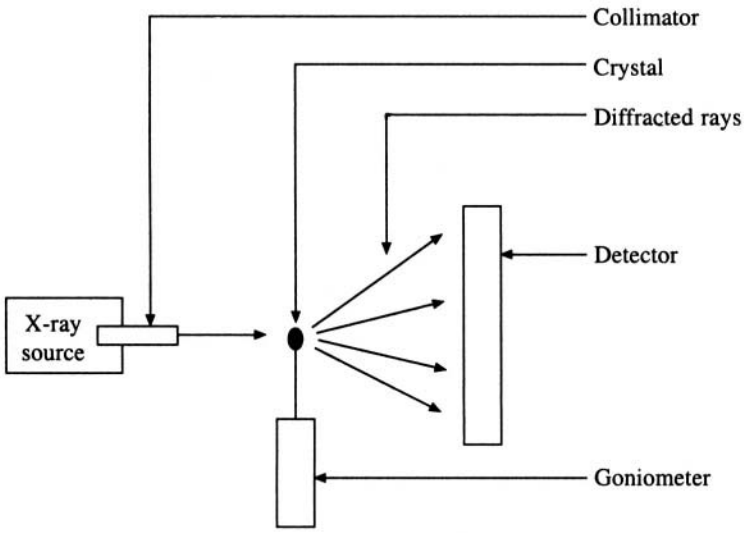


Fig. 7.13 A schematic sketch of the experimental setup for x-ray diffraction

The X-ray source is usually a sealed tube in which electrons are accelerated from one end and allowed to impinge at the other end on a metal target, usually copper or molybdenum for biologically relevant samples. This produces X-rays of wavelength 1.5418 Å (for Cu $K\alpha$) and 0.7107 Å (for Mo $K\alpha$). This process also produces a tremendous amount of heat and the tube has to be kept cool by circulating chilled water or by other means, such as rapidly rotating the metal target.

The collimated beam of X-rays then fall on the crystal mounted on a goniometer. This is a device

that allows one to position the crystal in different orientations in order to bring different sets of planes into the reflecting position.

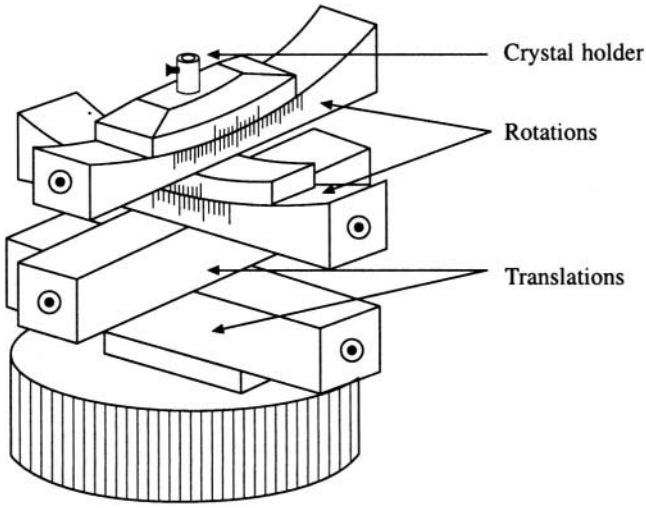


Fig. 7.14 A schematic diagram of a goniometer head, indicating the movements possible to centre the crystal in the X-ray beam

The detector of the diffracted X-rays may be a film, or a radiation counter. The first measurements of X-ray intensities used in the solution of the structure of NaCl, were made with an ionization chamber. This however allows the measurement of only one reflection at a time, since the crystal and the detector have to be set at different angles for each reflection. Films allow the recording of several reflections at a time, but suffer from the requirement of long exposure times, as well as imprecision in the conversion of the blackening of the film into a numerical value of intensity. Modern devices such as multi-wire proportional counters or imaging plates offer the advantages of both films and of counters and are now almost exclusively used in the collection of data from crystals of biological macromolecules.

The outcome of the data collection process is a list of angles, indicating the orientation of the crystal and the detector and the intensity of the reflection from the oriented lattice plane at this position. Instead of the angles, it is usual to identify the set of planes that give rise to each diffracted spot by the Miller indices *hkl* and associate with each value of *hkl* the appropriate value of the intensity. From the list of intensities, the amplitudes of each reflected wave is extracted by the simple process of taking the square root, since the measured intensity of a wave is the square of its amplitude. These amplitudes, represented as $\boldsymbol{F_{hkl}}$, are now used in the analysis of the structure of the molecule that makes up the crystal.

7.8 Structure Solution

The structure of the molecule is obtained by a Fourier transform of the observed amplitudes $\boldsymbol{F_{hkl}}$. This statement, which forms the core of the X-ray crystallographic method, is amplified and explained below. In order to first understand the mathematical process of Fourier transformation, consider the periodic function shown in Figure 7.15. Fourier theory shows that such a function can be represented as a sum of a series of simple waves. Note however that it is not only necessary to know the amplitude (and wavelength) of the simple waves but also to know the relative positions, or phases of the waves in order to reconstruct the original periodic function.

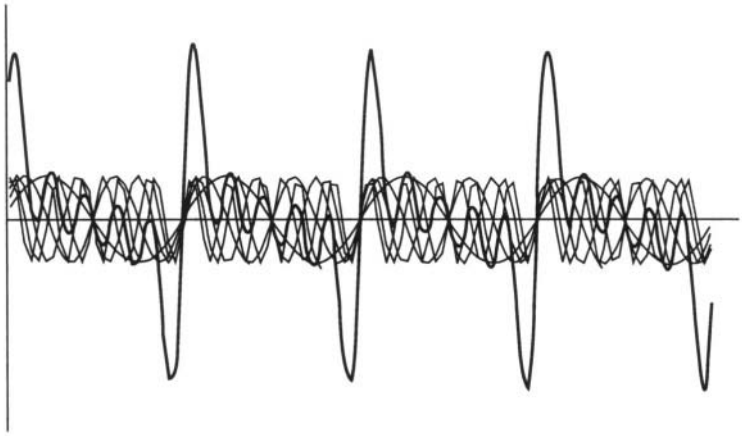


Fig. 7.15 The Fourier decomposition of a complex periodic function (thick line) into its sine and cosine components (thin lines). As more component waves at different frequencies (wavelengths) are added, the resulting wave approaches a square wave more and more closely. The Fourier decomposition of a square wave itself would result in components of all possible frequencies.

A crystal which has the unit cell, with its complex contents, repeated periodically, can be considered to be made up of simple planes of electron density arranged in three dimensions. This periodic function can be represented as series of simple waves. The amplitudes of these simple waves are measured, i.e. F_{hkl}. In order to reconstruct the crystal as a three-dimensional superposition of these simple waves of electron density, it is necessary to know the phases. Once the phases are known the mathematical operation of Fourier transformation, which corresponds to the physical superposition of the simple waves, will yield a picture of the distribution of electrons within the unit cell of the crystal. In other words, we obtain a picture of the structure of the molecules in the unit cell. The problem of X-ray crystallography now translates into one of obtaining the phase information. This is called the phase problem in crystallography since, in the X-ray diffraction experiments, the phases cannot be measured directly and the phases have to be determined by other indirect means. Some of these methods are described below.

7.8.1 Heavy atom or multiple isomorphous replacement (MIR) method

This method requires that the crystal of the sample be grown in at least two ways, once in the presence of a 'heavy', usually metal, atom and once in its absence. An additional, rather stringent, requirement is that in both cases the crystals are 'isomorphous', i.e. they have the same unit cell dimensions and crystal symmetry. In such a situation the only difference in the arrangement of the atoms in the crystal would be the presence or absence of the heavy atom. Whether a particular atomic species is called 'heavy' or not would depend on the contribution it makes to the total electron density and hence to the X-ray diffraction intensities. Thus for small organic compounds of molecular weight about 1000, cobalt, copper, iron and bromine would qualify as heavy atoms. For proteins, crystals may have to be grown in the presence and absence of platinum, samarium, mercury, etc. The heavy atom method depends on the assumption that, since a large portion of the scattering matter in the unit cell is concentrated into a single atom, a knowledge of the position of this atom and hence its contribution to the phases, would lead to a knowledge of the phases due to all the atoms in the unit cell, at least to a very good approximation. The position of the heavy atom itself is usually arrived at

by a special type of Fourier analysis called the 'Patterson synthesis'. Once the position of the heavy atom is known, the phases can be calculated. On the basis of the above stated assumption, they are the phases due to all the atoms and a Fourier transform will give the approximate positions of the atoms. Usually, especially for macromolecules like proteins, there are ambiguities in assigning the phases, which can be overcome only by growing many crystals by substituting different heavy atoms and collecting and analysing the X-ray data for each crystal. This method is therefore usually called the 'Multiple Isomorphous Replacement' method.

7.8.2 Direct methods

H. Hauptman and J. Karle pioneered a statistical analysis of the X-ray data to obtain the phases. It was realised that the values of the phases have to be such that the electron density obtained has to correspond to physical reality, i.e. there can be no 'negative' electron density in the crystal (the presence of an electron is considered as 'positive' electron density) and the density should correspond to separate or bonded atoms. In addition the phases for some of the reflections called semi-invariants can be restricted to a few values depending on the space group. These facts can be used iteratively in a computer program to generate the phases for all the reflections. A Fourier transform then gives the structure. The theory is very well developed for crystals with less than about 50 atoms in the unit cell and the technique is routinely used to determine small molecule structures. For larger molecules such as proteins, the method is still not fully successful.

7.8.3 Molecular replacement

Pioneered by M.G. Rossmann, the molecular replacement technique is used when the unknown structure can be assumed to be similar to an already known structure. The problem then becomes one of first finding the orientation and position of the model molecule in the unknown unit cell, and next of finding the differences between the model and the actual structure. The first part of the problem is usually solved by means of a search procedure. The computer, in effect, generates the F_{hkl} for each position and orientation of the model in the unit cell and compares these calculated reflections with those actually observed in the experiment. The position and orientation corresponding to the best match is then taken up for the second part of the structure solution procedure, i.e. the refinement.

7.8.4 The anomalous scattering technique

In this method the wavelength of the incident X-ray beam is adjusted so that the scattered intensities would be different when measured from opposite sides of the crystals. Normally the two sets of intensities would be the same, but in particular instances when the wavelength is close to the so-called absorption edge of one of the atomic species in the crystal, the intensities are different. This difference is then exploited to obtain the phases. Since the wavelength has to be tuned to match the element, this method became widely used after the advent of powerful, tunable (and very expensive!) sources of X-rays called synchrotron radiation sources.

7.9 Refinement of the Structure

Once the approximate value of the phases are determined the next step is to refine them. This usually more conveniently done by performing the Fourier transform and refining the approximate positions of the atoms obtained by including other known data such as stereochemistry.

7.9.1 Least squares method

This is a well-known mathematical technique used to fit a model to the observed data. Small changes

are made in the model and with each set of changes, the expected data are calculated. The differences between the observed and calculated data are once again used to determine the changes in model in an iterative process that terminates when further changes do not lead to any further improvement in the calculated data. Since the differences can be either positive or negative, i.e. the calculated data points can either be more or less than the observed data points, the sum of the squares of the differences, which is always a positive quantity, is the term that is actually minimised in the least squares procedure. The whole procedure can be described in terms of a mathematical operation, with the changes to the model in each cycle being calculated automatically. The mathematical equations lend themselves to easy computerisation and the least squares method is widely used in many fields of science and engineering. As applied to crystallography, the function to be minimised is called the Residual or the *R* factor and is usually defined as

$$R = (\Sigma||F_0 - |F_c||)/(\Sigma|F_0|)$$

where F_c is the calculated structure amplitude for a particular reflection *hkl, F* is the observed structure amplitude for the same reflection and the summation is over all the reflections. Note that only absolute differences between the calculated and observed data are taken into consideration, in keeping with the idea that the actual sign of the displacement is not relevant to the least squares technique. The calculation of the F_c involves a Fourier transform of the model and it is performed in each cycle after shifting the atoms by the (usually small) amounts indicated in the previous cycle. Though the amount of computation required is quite large, modern technology allows the complete refinement of structure of molecular weight of about 1000 in less than one day. In order to work, the least squares technique requires that the initial model be close (< 0.5 Å) to the actual final model. Furthermore a very large number of observations are required for reliable results. As a rule of thumb, between seven and ten observed data points are required for every parameter refined. For example, a molecule containing 20 atoms will have three positional parameters (*x*, y, z) and six parameters corresponding to the thermal vibration, a total of nine parameters for each atom. To ensure accurate refinement therefore, more than 1500 reflections will have to be measured. In the case of biological macromolecules, this condition is seldom satisfied. Owing to various factors, including high solvent content of the crystals, the molecules are not very well ordered and reflection planes with spacing of less than about 2.0 Å are usually not sufficiently regularly spaced to give measurable diffracted intensities. In other words the limit of resolution of the data in such cases is usually not beyond 2.0 Å. This means that the observation-to-parameter ratio is not good enough to allow refinement using the standard least squares techniques.

7.9.2 Constrained-restrained refinement

There are two ways in which the observation-to-parameter ratio can be improved without having to collect additional X-ray data. The first is to increase the number of observations by including known chemical information. For example, in the extreme case, the whole molecule can be treated as a single rigid unit. The only variables are then the orientation and the position of the molecule, a total of six parameters. This is called constrained refinement. In the intermediate cases various parts of the molecule can be considered rigid units, which are then linked to each other by flexible bonds. Each such unit will have six parameters, as compared to the three parameters for each atom of the unit. A typical example is the case of the DNA molecule (Figure 7.16) where each base can be considered to be a rigid planar unit. Likewise the sugar is a rigid unit, as also the phosphate group. Obviously the total number of parameters to be refined would reduce drastically, leading to a much better observation-to-parameter ratio. Another method of improving the ratio based on available stereochemical information

is called the restrained refinement technique, where the distances between the various atoms in the molecule are known. For example, the distance between two carbon atoms joined by a single bond cannot be very different from 1.54 Å. Similarly the angle at a sp^3 hybridised carbon atom has to be close to 109.5°. This chemical information is treated as additional observation points and for the same number of parameters, the ratio improves. The bond length and angle information, together with other stereochemical and physicochemical information such as van der Waals contacts, hydrogen bonds, steric hindrances and intra-molecular electrostatic forces can also be programmed as an energy function which is to be minimised. This is then added to the residual function mentioned earlier and the whole is minimised by least squares or some other optimisation technique.

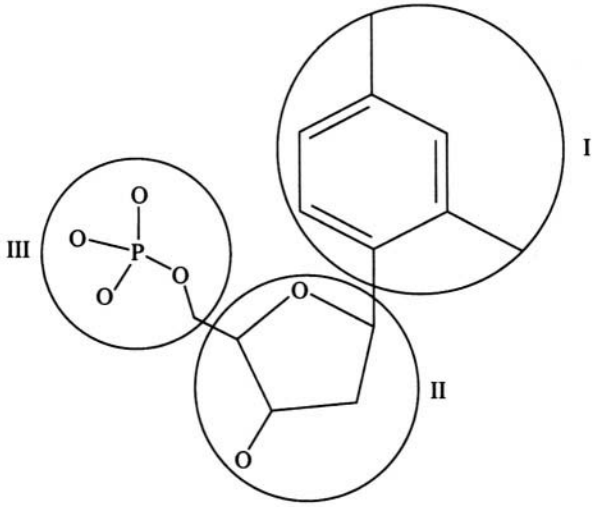


Fig. 7.16 The rigid groups in a nucleotide unit. I: atoms of the base are considered as one rigid unit. II: atoms of the ribose moeity. III: atoms of the phosphate group.

7.9.3 Computer graphics in refinement

The advent of high-resolution computer graphics has made possible the development of several new techniques of refinement and of utilising known chemical and biological information in the refinement process. Today much of the crystallographer's creativity is expressed at the computer graphics terminal. As mentioned earlier, once the approximate phases have been obtained, a Fourier transform using the amplitudes F_{hkl} as the Fourier coefficients produces a map of the electron density in the unit cell (Figure 7.17). The model is superposed on this map and wherever the fit of the model to the electron density is not good enough, adjustments are made to improve the fit. After a cycle of fitting, or changing various parts of the model, the map can be recalculated using the changed positions of the atoms in the model to obtain the new set of phases. This process iteratively brings the model closer and closer into a perfect fit with electron density. Simultaneously a constrained or restrained least squares refinement procedure is also carried out and the conventional Residual or *R* factor is also used as a check on the correctness of the structure.

At the end of the refinement process we obtain a set of coordinates for each of the atoms of the molecule and another number (or set of numbers) representing the thermal vibrations of the atom. These can be subjected to further geometrical analyses to understand the chemistry or biology of the molecule.

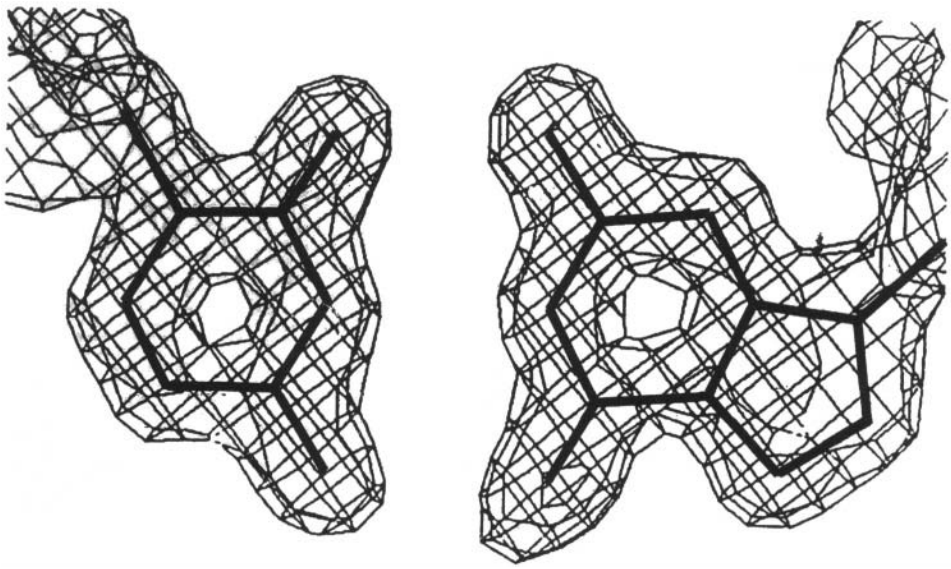


Fig. 7.17 Electron density net corresponding to a G.C. base pair in the crystal structure of a DNA fragment.

7.10 Note on the Resolution of an X-ray Structure

The number of X-ray reflections that may be obtained from a crystal depends on the degree of order present in it. A greater degree of order implies that lattice planes with smaller inter-planar spacing are regularly arranged in the crystal and that X-ray reflections from these planes can also be measured. This implies that in the final structure, features at smaller distances can be resolved (i.e. seen as separate entities). The terminology of the resolution can be somewhat confusing. Since a higher resolution means smaller separation of the observable features, a resolution of 6 Å, is lower than a resolution of, say, 2 Å. The larger the number the lower the resolution and vice versa. The resolution can also be expressed as the angular limit at which the intensity data can be observed. By Bragg's law

$$\sin \theta = n\lambda/2d$$

In other words, the lower the inter-planar spacing d (i.e. the greater the resolution) the higher the angle at which the diffracted beam corresponding to that set of planes is observed. The terminology in this case is more straightforward since $\sin\theta$ varies directly as the resolution. If we normalize with respect to the wavelength of the incident radiation, the resolution may be expressed as $\sin\theta/\lambda$ (or 1 2d). Thus, the highest resolution achievable using radiation of a particular wavelength is $d = \lambda/2$ Å

8

NMR Spectroscopy

8.1 Introduction

Nuclear Magnetic Resonance Spectroscopy is one of the most important experimental techniques currently in use in biophysics. Until the early 1970's NMR was of interest mainly to physicists and chemists. Physicists use NMR Spectroscopy to study the nature and interactions of the atomic nucleus. Chemists use NMR Spectroscopy to determine chemical structure of small molecules. With the development of techniques such as Fourier transform NMR, and two-dimensional and three-dimensional NMR, biochemists and biophysicists use the method to determine the structures of large molecules such as proteins and also to study the dynamic behaviour of such molecules. Magnetic Resonance Imaging is becoming increasingly popular in medicine as a means to obtain accurate three-dimensional images of the interior organs of the human body in an essentially safe and non-invasive manner.

NMR is a rapidly growing field and already it is so vast that only a few of the specialised techniques can be explained in the present book. In this chapter we first explain the basic principles of the NMR technique. NMR theory is then dealt with in somewhat greater detail including a description of the parameters measured in the different types of experiments. Finally we discuss the applications of NMR in chemistry, biology and medicine.

8.2 Basic Principles of NMR

NMR is a spectroscopic technique. In common with other spectroscopic techniques, the NMR experiment consists of the measurement of the intensity of absorption (or emission) of electromagnetic radiation at each frequency over a specific range. In NMR this range is the radio-frequency range, i.e. from a few MHz to a few hundred MHz. At this frequency, the energy is absorbed or emitted when the nuclei of the atoms go from one quantized spin state to another. Spin is a property possessed by only certain atomic nuclei such as the hydrogen nucleus *H* or the phosphorous isotope $^{31}\mathbf{P}$ (Figure 8.1). A crude classical analogy to this property of spin is the rotation of a body about an axis that passes through

its centre. Since the nucleus has a charge, the spinning motion will give rise to a magnetic field, and the spinning nucleus can be thought of as a small bar magnet. The axis of the bar magnet will be parallel to the axis of spin. When this nucleus is placed in a strong external magnetic field, then, by the laws of quantum mechanics and due to the interaction of the two magnetic fields, the axis of spin will orient itself in one of two possible orientations (Figure 8.2). Transitions between these two states can occur when another external oscillating magnetic field supplies sufficient energy for the spin system, i.e. the nucleus to go from the state or orientation of lower energy to the state of higher energy. The difference in energy of the two states corresponds to the energy of a radio-frequency field and hence a nuclear spin, placed in a static magnetic field, must be excited by radio-frequency electromagnetic radiation in order that the transition takes place. The energy absorbed from the radio-frequency field can be detected and this is the fundamental parameter measured in an NMR experiment. The frequency at which the transition occurs is called the resonance frequency The resonance frequency of a particular type of nucleus, e.g. ^{1}**H,** will be different when it is placed in different chemical environments. For example the ^{1}**H** in a methylene group will have a distinctly different resonance frequency from one in a methyl group. By changing the frequency of the exciting radio-frequency field and measuring the intensity of the absorption at each frequency, it is possible to detect the nuclei and in each case identify their respective environments. It is this feature which is useful in identifying various chemical groups in molecules. If the magnetic field in which the spins are placed is very high, subtler effects become noticeable, i.e. the resolution of the spectrum increases. In this way it is possible to detect even conformational changes in the molecule.

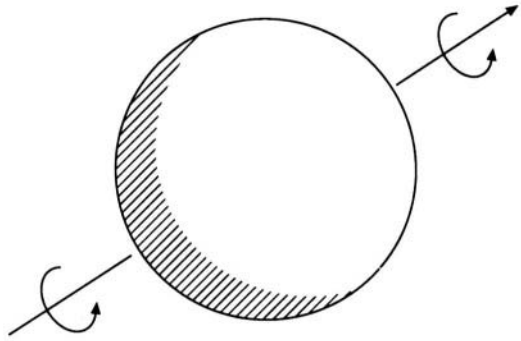


Fig. 8.1 The spin of a nucleus

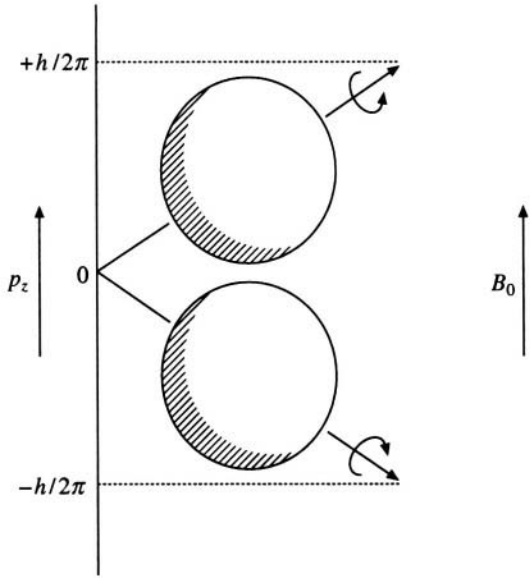


Fig. 8.2 The two possible orientations of nuclear spin in a magnetic field

Continuously changing the frequency of the radio field and measuring resonance is a tedious and time-consuming task. So much so, in the case of large molecules, where the changes in intensity are very small, formidably long data collection time would be required. Fourier transform NMR and multi-dimensional NMR have reduced the experimental time required immensely and have made it possible to use NMR to study large biological molecules and complexes.

Another aspect of the NMR experiment is used in medical imaging. After a nucleus makes a transition from a lower energy state to a higher energy state, it has a tendency to 'relax' back to the lower energy state by transferring the extra energy it has acquired either to another spin nearby, or to

the other atoms and molecules, i.e. the 'lattice'. The time required for this relaxation will depend upon the nature of the lattice. In a living body, different tissues have different relaxation times. A measurement of the relaxation time of the spin nuclei in different parts of the body can be made. These measurements can then be displayed as a picture of the interior of the body, enabling medical practitioners to detect malfunctions or tumours etc. in the various organs.

8.3 NMR Theory and Experiment

Rigorous NMR theory requires a knowledge of quantum mechanics and will not be attempted here. Some of the results obtained will be stated without proof. According to quantum mechanics the angular momentum of atomic nuclei or, in other words, the spin of an atomic nucleus can only have certain discrete values characterised by the spin quantum number I. The value of I for any nucleus can either be an integer or a half-integer (1/2,3/2,5/2 …) and depends on the number of protons (the atomic number and the number of neutrons present. (The sum of the neutrons and protons is the mass number. Table 8.1 gives the values of the spin for various possible combinations. For example the nuclei ^{1}H, ^{13}C, ^{31}P, and ^{19}F all have odd mass numbers and therefore a half-integral spin value. (In these particular cases the value is 1/2). The nuclei ^{12}C and ^{16}O have even atomic numbers (6 and 8 respectively) and even mass numbers and the spin is zero. Nuclei of deuterium, i.e. ^{2}H, and ^{14}N have odd atomic numbers (1 and 7 respectively) and even mass numbers. The value of the spin is 1. In chemistry and biology, the most useful nuclei are those with spin 1/2 and the following discussions will be restricted to such nuclei.

Table 8.1 Relation between the nucleon numbers and spin. Only nuclei with half-integral or integral spins are NMR-active

Atomic number	Mass number	Spin quantum number
Even or odd	Odd	Half integral
Even	Even	Zero
Odd	Even	Integral

When placed in a strong, uniform magnetic field, again according to quantum mechanics, the spin angular momentum vector can take up one of two positions (Figure 8.2) such that the component of the angular momentum along the direction of the applied magnetic field $\boldsymbol{B}_0$ is

$$p_z = \pm 1/2(h/2\pi)$$

where h is Planck's constant. The z component of the magnetic moment of such a nucleus is also quantized and is given by

$$\mu_z = \gamma p_z = \pm 1/2(h/2\pi)\gamma$$

where γ is a parameter that depends on the particular nucleus and is called the magnetogyric (or gyromagnetic) ratio. The energy associated with these two positions due to the interaction of the magnetic moment with the applied magnetic field is

$$E = -\mu_z B_0 = \pm 1/2(h/2\pi)\gamma B_0$$

Thus the difference in energy between the two states is

$$\Delta E = \gamma(h/2\pi)B_0$$

Therefore, to induce the spin to make the transition from the lower energy state to the higher energy state, the applied oscillating magnetic field, i.e. the radio-frequency field should supply this energy. This implies that the frequency of oscillation should be ν where

$$\Delta E = h\nu_0$$

and

$$\nu_0 = \gamma B_0/2\pi$$

If we write $\omega_0 = 2\pi\nu_0$, where ω_0 is the angular frequency, then

$$\omega_0 = \gamma B_0$$

This is called the resonance condition and is the basic NMR equation.

Actual NMR experiments, of course, do not deal with single nuclei and single spins. The sample will contain a large number of nuclei, of the order of Avogadro's number ($\sim 10^{23}$). When the sample is placed in the static magnetic field, all the nuclei will not take up the lower energy state. Instead the ratio of the number of spins in the lower energy state to the number in the higher energy state is given by the Boltzmann formula

$$\exp(-\Delta E/kT)$$

where k is the Boltzmann constant and T the absolute temperature of the sample. Since

$$\Delta E = \gamma(h/2\pi)B_0$$

the ratio is

$$\exp(-\gamma(h/2\pi)B_0/kT).$$

This means that at room temperature and at medium magnetic fields, the energy difference between the two states is such that the population of spins in them differs by less than 1 part in ten thousand. This implies that the signal is very weak, since only a relatively few nuclei will absorb energy and make the transition. Note that if the applied static field is increased, the energy difference increases, leading to a larger population difference and hence a stronger signal.

8.4 Classical Description of NMR

The NMR condition

$$\omega_0 = \gamma B_0$$

may also be derived from considerations of classical mechanics, as follows. If we consider the nucleus to be analogous to a spinning bar magnet, classical mechanics tells us that such a system placed in a magnetic field will precess about the direction of the external field with the so called Larmor frequency given by

$$\omega_0 = \gamma B_0$$

The large numbers of nuclei in a typical experimental sample will all precess about $\boldsymbol{B}_0$ (Figure 8.3). This would mean that, while the net component of the magnetic moment in the xy plane is zero, there will be a net magnetisation along the z-axis (presuming $\boldsymbol{B}_0$ to be along z, as in Figure 8.3). In order to perform the NMR experiment, it becomes necessary to tilt the total magnetisation towards or onto the xy plane as illustrated in Figure 8.4. (This is equivalent to inducing a transition of many nuclei, from the lower energy quantum state to the higher energy state). To understand this better, we may

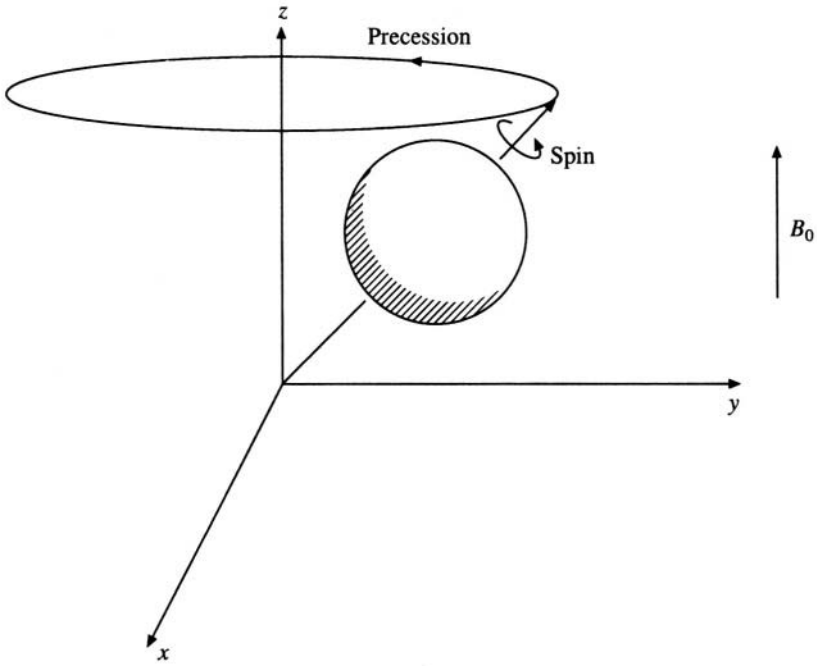


Fig. 8.3 Precession of a spinning nucleus. The precession axis is the z axis.

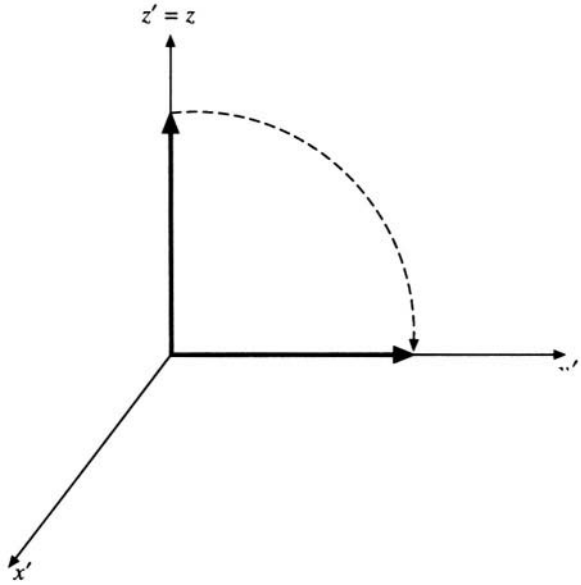


Fig. 8.4 Tilting of the total magnetisation on to the xy plane

change our reference axial system from the static xyz system in Figure 8.3 to the rotating frame of reference given by $x'y'z'$. This reference frame is assumed to rotate about the z-axis with the same angular frequency as the Larmor frequency ω_0. Thus in this frame, the nuclear magnets do not precess

and the apparent static magnetic field is zero. Now if a static magnetic field B_1 is applied along the x' direction in this field, then, just as the field B_0 caused a precession in the stationary reference frame, so will the field B_1 cause the nuclear magnets to precess about the x' axis, and the total bulk magnetisation will rotate about the x' axis. Of course the field B_1 is not actually a static field, but only appears so in the rotating frame. In fact B_1 is the radio-frequency field (of frequency ω_0 = Larmor frequency) which is applied in addition to B_0 in the NMR experiment. If the field B_1 is applied for a time *t,* then the magnetisation will rotate through an angle given by $\theta = \gamma B_1 t$. If the duration of the pulse of radio-frequency electromagnetic radiation is such that the magnetisation rotates by 90° onto the *xy* plane, then such a pulse is called a 90° pulse. In the same way a 180° pulse will reverse the direction of the magnetisation. The pulse produces a net component of the magnetisation in *xy* plane also. This component then produces an effect upon the detection coil in the NMR spectrometer, which is amplified and displayed as the NMR spectrum.

There are two main methods by which the NMR experiment is performed, viz. the continuous wave method and the Fourier transform method. In the continuous wave experiment the radio-frequency field is continuously applied to the sample, and the magnetic field is changed continuously in strength such that all the nuclei in the sample are able to come into resonance. Alternatively the magnetic field may be kept fixed and the oscillating radio frequency field may be varied. Either way, the continuous wave method, has been almost entirely replaced by the Fourier transform NMR (FTNMR).

In FTNMR the radio-frequency stimulus is applied as a pulse which may last typically for about $200\mu s$. The pulse is sufficiently strong and is constituted of a sufficiently large number of frequencies to excite all the resonances in the sample. As a matter of fact what is usually applied is a square pulse. As explained in a previous chapter (Chapter 7), according to Fourier transform theory, such a square pulse is constituted of all possible frequencies. The response to such an excitation would be an NMR signal from each of the spins in the sample. The composite would be another complicated wave which may be Fourier transformed to extract the NMR spectrum. In order to improve the detection of the signal, or, in other words, the signal to noise ratio, this process of giving the pulse and measuring the response is repeated several times, before the Fourier transform is taken. If the response is measured *N* times and added, the signal will increase *N* times while the background noise, which is random will increase only $\sqrt{N}$ times. The expression free induction decay or FID is used to describe the response on the application of the radio-frequency pulse – it describes the decay of the induced signal arising from the free precession of the nuclei in the field B_0

The nuclei in the sample will be present in several different environments, and the resulting FED will consist not only of several different frequencies, but also each different frequency will decay at a different rate. The Fourier transform of the FID therefore contains information not only about the resonance frequency of each nucleus but also about the decay of the resonance and hence the rate and mode of energy transfer from the nucleus to the environment.

8.5 NMR Parameters

In general, the following are the measurable NMR parameters: (1) chemical shift (2) intensity (3) line-width (4) spin-spin coupling constants.

8.5.1 Chemical shift

The effect of the external applied magnetic field on each of the protons in a molecule will not, in general, be the same. This is because, owing to the electron currents surrounding each nucleus, an additional small field is also created just around the nucleus. The total effective field at each nucleus

will thus depend on the neighbouring atoms and chemical groups. Since the magnetic field is different at each chemically distinct nucleus, the frequency at which the resonance occurs will also be different at each nucleus. Consider for example a molecule such as ethanol. A schematic sketch of its ^1H NMR spectrum at 100 MHz (i.e. B_0, the applied field strength, corresponding to a resonance frequency of 100 MHz for a free H nucleus) is given in Figure 8.5. The frequency at which the hydroxyl (OH) proton resonates is different from the one at which the methylene (CH_2) protons resonate which is again different from the one at which the methyl (CH_3) protons resonate. Such a shift in the position of the resonance peak in the spectrum due to the chemical environment is called the chemical shift.

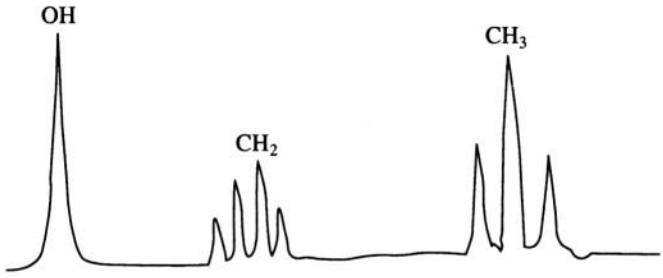


Fig. 8.5. The NMR spectrum of ethanol (schematic). The splitting of the peaks has been exaggerated.

The chemical shift is always measured with respect to some arbitrary chosen reference point in the spectrum. Usually an inert compound with a well-known resonance position, not much affected by the molecular environment, is used as the standard. The standard may be introduced either internally (i.e. mixed with the sample) or externally (in a small tube placed alongside the sample tube). Chemical shifts are expressed in the dimensionless units of parts per million (ppm) and are normalised to be independent of the applied magnetic field. The chemical shift

$$\delta = (\nu_s - \nu_R)/\nu_R \times 10^6$$

where υ_S is the resonance frequency of the sample, and ν_R is the resonance frequency of the reference. By convention, when displaying the NMR spectrum, the resonance frequency of the standard is placed at the right-most end of the scale at 0 ppm, and the scale, in ppm, increases to the left, in decreasing order of frequency. This corresponds to the fact that the chemical shift is actually produced by a shielding and therefore a reduction of the applied magnetic field by the electrons in the neighbourhood of the resonating nucleus. Table 8.2 indicates the typical chemical shifts for some important chemical groups.

8.5.2 Intensity

The signal strength or intensity of the resonance peak is a measure of the number of nuclei that contribute towards it. Signal intensity is obtained by measuring the area under the peak. To be mathematically rigorous, it is necessary to integrate each peak from $-\infty$ to $+\infty$. In practice this is not possible and one has to be content with measuring the area under those parts of the peak which are accessible. This usually accounts for only about 90% of the total area, and hence there is always an error of ±10% in the measurement of the signal strength.

Apart from the number of nuclei resonating at the given frequency, signal intensity is also influenced by the effects of saturation, the nuclear Overhauser effect, the effects of a finite pulse width and other experimental limitations. If these factors are taken into consideration and if appropriate corrections

Table 8.2. Typical ^{1}H chemical shifts (in ppm from tetramethysilane or TMS. Where a range is given, the position of the peak is dependent on the attached groups as well as the solvent, pH, etc.)

Chemical group	Chemical shift δ
$-CH_3$	0.9
$-CH_2$	1.4
–CH	1.5
$-C_6H_6$	7.3
–CHO	9.7
–OH	0.5-4.0
$-NH_2$	2.0-5.0
–NH	2.0-5.0
–SH	1.0-2.0

are applied, then accurate measurements of relative concentrations of the different nuclei can be made. If absolute concentrations need to be measured, a known quantity of a standard sample can be used to calibrate the spectrum and measurements can then be made. This technique is especially useful in biomedical applications.

8.5.3 Line width

The width of each spectral line is related to the relaxation processes in the sample. There are primarily two ways in which the additional energy gained by the nucleus on transition can be released. The energy can either be transferred to a neighbouring spin or it can be transferred to the neighbouring lattice. A time constant is associated with each and referred to as T_2 and T_1 respectively. Line width is measured at half the signal height in Hz and is represented as $\Delta\upsilon_{1/2}$. The relationship to so-called spin-spin relaxation time T_2 is

$$1/T_2 = \Delta\nu_{1/2}$$

in the ideal situation where only the relaxation contributes to the line width.

8.5.4 Relaxation Parameters

As explained earlier, when a sample is placed in a magnetic field, it becomes magnetised. The direction of the magnetisation M_0 is the same as that of the applied magnetic field, being conventionally taken as the z-axis. The components of magnetisation along the other two directions, i.e. M_{xy} is zero. This is the situation in the equilibrium condition. When a radio frequency pulse is applied, the magnetisation tilts over onto the *xy* plane and as the perturbing field is removed, the magnetisation returns to the equilibrium situation of $M_z = M_z$ and $M_{xy} = 0$. This happens by the relaxation processes and lasts over a period of time ranging from milliseconds to a few seconds. As mentioned earlier, there are two main relaxation processes, spin-lattice relaxation and spin-spin relaxation.

Spin-lattice relaxation is responsible for the return of the z component M_z to its equilibrium value. It is also called longitudinal relaxation. The lattice referred to in the term is the environment and all the nuclear dipoles surrounding the particular spin, whether they belong to the same molecule or to other molecules. In solid state NMR, the word lattice may refer to the crystal lattice. In solution phase NMR it simply means the neighbourhood of each spin. The return of the nuclear spins to thermal

equilibrium, i.e. to the equilibrium of the lattice is characterised by a time constant referred to as T_1. A measurement of T_1 is frequently an important part of the NMR experiment, and is required for the following reasons: (a) To select the most suitable radio-frequency pulse interval, the T_1 values of the various resonances need to be known. (b) Knowledge of T_1 values helps in determining the rates of chemical reactions through NMR. (c) T_1 provides information about the molecular structure. (d) In modern medicine T_1 is the most useful parameter used in NMR imaging. Frequently the image of the tissue seen is nothing but a map of the values of T_1 at different positions within the tissue. T_1 may be measured by FTNMR by applying a sequence of radio-frequency pulses. There are several different pulse sequences that may be used. A commonly used sequence is the inversion recovery sequence in which a 180° pulse is first applied. (As already explained, a 180° pulse changes the angle of the bulk magnetisation vector by 180°). The magnetisation is allowed to relax back to its equilibrium value for a time interval τ and a 90°pulse is applied and the FID is collected soon after. This signal will contain information about the strength of M_z after time τ. If the process is repeated for different values of τ, the spin-lattice relaxation time constant T_1 may be obtained.

Spin-spin relaxation is responsible for the return of the M_{xy} component of the magnetisation to its equilibrium value of zero. This is also known as transverse relaxation and is characterised by a time constant T_2 known as the spin-spin or transverse relaxation time. Consider Figure 8.6. Immediately after the application of the radio-frequency pulse, the net magnetisation is given by some maximum value (Figure 8.6(a)). If there is only a single spin in the sample resonating at precisely one frequency, the magnetisation (as considered in the rotating frame of reference) remains static and of constant magnitude. However, if there is spread of frequencies, different nuclei will precess at different frequencies. In the rotating frame of reference the magnetisation vectors for the spins will fan out, resulting in a reduction of the net magnetisation (Figures 8.6 (b) and (c)). Finally M_{xy} becomes zero, the equilibrium value (Figure 8.6(d)). Since T_2, the time constant characterising this process depends on the spread in frequencies, the width of the resonance peak is a measure of T_2.

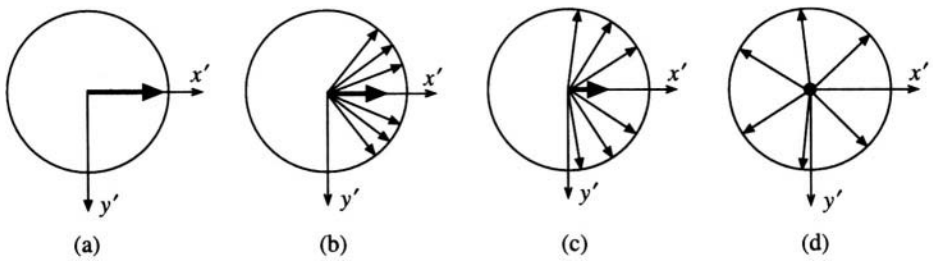


Fig. 8.6 Spin-spin relaxation—the return of M_{xy} to its equilibrium value.

Spin-spin relaxation processes are therefore those that cause line broadening. One component of this is the dipole-dipole interaction, which also causes spin-lattice relaxation, and therefore T_2 cannot be longer than T_1. Additional contributions to T_2 come from the z components of the small perturbing fields of the neighbouring spins. T_2 measurements provide information about molecular structure. They can also be used to determine rates of molecular diffusion. Information about T_2 helps to design pulse sequences that may simplify the spectra and enhance the resolution.

8.5.5 Spin-Spin Coupling

This is different from the spin-spin relaxation process explained above. Resonances in the NMR spectrum often do not occur as a single peak but are observed to be split into doublets, triplets,

quadruplets or higher multiplets. The splitting is due to the neighbouring spins that are transmitted via the electrons of the covalent bonds between the atoms. Spin-spin coupling is thus a 'through-bond' coupling and occurs only due to nuclei within the molecule. Splitting of the peak can be caused only by nuclei with spin $I \neq 0$. The magnitude of splitting is designated by the spin-spin coupling constant *J*. It is measured in Hz and is independent of the applied magnetic field strength. Figure 8.7(a) is the spectrum of acetaldehyde (CH_3CHO) in which the CH_3 signal is split into a doublet due to spin-spin coupling of the protons with the neighbouring CHO proton through the three bonds between them (Figure 8.7(b)). The CHO resonance itself is split into a quadruplet owing to coupling with the three methyl protons. A complete theoretical treatment of spin-spin coupling is beyond the scope of this book. We substitute such a treatment by the following discussion.

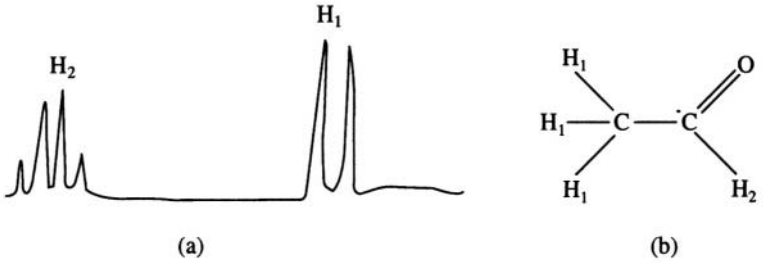


Fig. 8.7 (a) The NMR spectrum of acetaldehyde showing the splitting of the peaks (schematic) (b) Chemical structure of acetaldehyde.

Spin-spin coupling can be first order or second order. In the case of first order splitting the difference in frequency between the resonance peaks involved in the coupling is much greater than the coupling constant *J*. In the case of second order spectra this is not so. Obviously second order spectra will be much more difficult to interpret than first order spectra. First order multiplicities can be predicted by the following rules. (a) A singlet is a peak undisturbed by the presence of neighbouring nuclei. (b) Each neighbouring nucleus with spin $I \neq 0$ will split the peak into $(2I+1)$ components. If there are *n* equivalent nuclei, then the peak is split into $(2nI + 1)$ separate peaks. In Figure 8.7, the CHO peak is split by the three equivalent methyl protons (spin 1/2) into $(2 \times 3 \times (1/2) + 1) = 4$ components. (c) The components of a multiplet are evenly spaced and the spacing is a measure of the coupling constant *J,* in Hz. (d) The intensities of the components of the multiplet are proportional to the coefficients of binomial expansion. In the case of spin 1/2 nuclei, for example, the possible multiplets and the ratio of the peak heights are singlet, doublet 1:1; triplet 1:2:1; quartet 1:3:3:1; etc. (e) Since it is only required that the neighbouring nucleus possesses a spin $I \neq 0$, coupling between ^{13}C and ^{1}H (i.e. heteronuclear coupling) can occur as well as homonuclear coupling. However the relative abundance of ^{13}C is so small that such effects are usually not noticeable.

Coupling constants are affected by the following factors, among others; (a) dihedral angle between the nuclei (b) the electronegativity of the substituents on the moeity (c) valence angles (d) bond lengths. Figure 8.8 illustrates these parameters for a H-C-C-H molecule.

An empirical equation called the Karplus equation gives the dependence of the coupling constant on the dihedral angle of the above system.

$$J = 8.5 \cos^2 \phi - 0.28 \quad \text{for } \phi = 0\text{-}90°$$

$$J = 9.5 \cos^2 \phi - 0.28 \quad \text{for } \phi = 90°\text{-}180°$$

Clearly, measurement of the coupling constants is a sensitive method of determining molecular conformation as will be illustrated later.

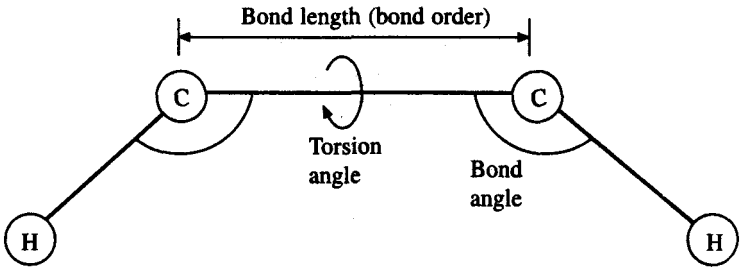


Fig. 8.8 The geometrical factors that affect the coupling constants for a H—C—H system

8.6 The Nuclear Overhauser Effect

The nuclear Overhauser effect or NOE illustrates the so-called 'double resonance' technique in which a second radio frequency field is applied to the sample in addition to the observing radio frequency field. The two fields may be denoted as H_2 and H_1, respectively. The second, strong, H_2 field is usually used to saturate particular resonances and is used in spin decoupling, saturation transfer or NOE experiments. H_2 is maintained at the resonance frequency of one resonance, to saturate it while another resonance is monitored by the non-saturating H_1 field. The monitored resonance may change in intensity if affected by the saturated resonance due to a change in the populations of the nuclear energy levels. Under normal conditions the relative populations of the spin states (or nuclear energy levels) is given by the Boltzmann distribution. If this distribution is altered by a saturation of some of the nearby resonances, and the consequent transfer of magnetisation by dipole-dipole coupling, it can lead to an enhancement of the observed signal. Strictly speaking a theoretical framework for NOE enhancement can be constructed only when the coupling is entirely a dipole-dipole coupling. Nevertheless it is possible to obtain an excellent indication of the degree of dipolar coupling between the nuclei by a measurement of the degree of NOE enhancement. The magnitude of dipolar coupling is strongly distance dependent and is proportional to $1/r^6$ where r is the distance between the nuclei. NOE is a 'through-space' effect and unlike spin-spin coupling, it does not need the interacting nuclei to be covalently linked to each other. Thus NOE is a powerful aid to the conformational analysis of molecules. A typical NOE experiment would consist of identifying all the resonances in the molecule and then saturating them one by one, each time measuring the enhancement of the intensity of the other resonances. In this way, it is possible for example to identify those portions of a polypeptide chain which may fold back to be close together in space though they are far apart in sequence.

8.7 NMR Applications in Chemistry

8.7.1 Chemical shift

The proton and ^{13}C chemical shifts in an NMR spectrum are a sensitive measure of the chemical environment of the nucleus. The shifts are influenced by a number of chemical factors, which are slightly different for protons and for ^{13}C nuclei. These are discussed separately below.

Table 8.2 gives the proton chemical shifts (in ppm from TMS) for some common chemical moieties. Some of the factors which influence the proton chemical shifts are: the state of hybridisation of the atom to which the proton is attached, van der Waals effects, anisotropy induced by the bond order of the neighbouring bonds or aromatic rings, solvent effects, effects of concentration and temperature, hydrogen bonding, contact shifts in organometallic compounds and chemical shift due to charged species.

The state of hybridisation of the atom (usually carbon atom) to which the proton is attached influences the amount of deshielding of the proton and hence the chemical shift. As the *s* character of the hybridisation increases, the resonance shifts more and more downfield (i.e. in the direction of increasing values of ppm or to the left of a spectrum as normally drawn). Thus a proton attached to a '*sp*' hybridised carbon will have its peak 5 to 8 ppm downfield of one attached to sp^3 hybridised carbon.

Non-bonded nuclei when pushed close to a proton distort the electronic shield around it, leading to an influence on the chemical shift.

Anisotropy effects on the chemical shift arise due to the fact that the electrons around neighbouring nuclei are asymmetrically distributed. This results in a variation of the chemical shift when the position of the resonating spin is changed with respect to the neighbouring nucleus. Thus in a molecule such as cyclohexane, protons which are equatorial to the ring will have different chemical shifts from those which are axial. Anisotropy effects are dependent on the order of the neighbouring bonds. Single bonds lead to less anisotropy than double or triple bonds. In aromatic rings, the circulation of π electrons around the ring results in a shielding of nucleus above and below the plane of the ring, leading to an upfield shift, and a deshielding of protons in the plane of the ring, leading to a downfield shift of the resonance. Van der Waals or anisotropy effects are also produced by the solvent in which the sample is dissolved.

Hydrogen bonding produces a downfield shift in the resonance of the protons involved in the hydrogen bonds, since such a bond is between electronegative atoms, which would withdraw the electron away from the proton leading to a deshielding.

8.7.2 Spin-spin coupling

Spin-spin coupling is dependent on the nature of the neighbouring nuclei and coupling constants carry information about the chemical structure of the molecule. It is however necessary to first of all distinguish between chemical equivalence of atoms and the magnetic equivalence of atoms. In the case of chemical equivalence the spins have identical chemical environments and thus identical chemical shifts. Magnetically equivalent nuclei will not only have the same chemical shift but will also possess the same coupling constant with every other nucleus in the molecule. Figure 8.9(a) shows tert-butanol in which the methyl hydrogens are both chemically and magnetically equivalent. In Figure 8.9(b) the methylene hydrogens of diflouroethylene are chemically equivalent but magnetically non-equivalent since each hydrogen is coupled to each of the two flourines with a different coupling constant.

Coupling constants in molecules can be either positive or negative in sign and up to a few tens of Hz in magnitude. Of the chemical factors affecting spin-spin coupling, the dihedral angle, the bond

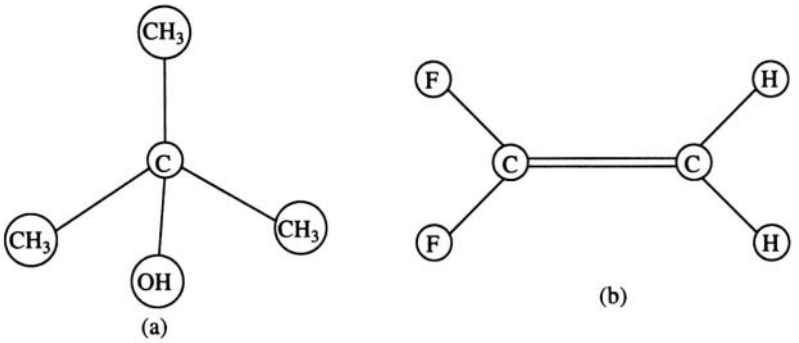


Fig. 8.9 Chemical and magnetic equivalence. (a) Tert-butanol. (b) Difluoroethylene

length and bond angle in a H-C-C-H system has already been discussed. Other factors are: (1) the electronegativity of the substituents, (2) valence angles and (3) bond lengths. Vicinal protons, i.e. protons such as those in the system H-C-C-H, show a decrease in the coupling constant if an electronegative substituent is attached to the carbon atoms. If more than one such substituent is present, the decrease is greater. In cyclic systems the steric disposition of the electronegative substituent also affects the coupling constant, being larger when it is equatorial and smaller when it is axial. The effect is the greatest when the substituent is trans-coplanar to one of coupling protons. For geminal protons, i.e. protons attached to the same carbon atom, the coupling is usually negative. An additional electronegative substituent decreases the values further (magnitudes increase, but the sign remains negative). Neighbouring π bonds such as those found in aromatic systems decrease coupling constants still further. Vicinal coupling constants are sensitive to the angles between the C–H and C–C bonds, as well as to the C–C bond lengths. In both cases the coupling constants decrease as the values of the angles or bond lengths increase. Long range spin-spin coupling extending over four or five bonds is also possible, through the magnitude is usually small, of the order of a few Hz.

8.7.3 ^{13}C NMR

Among the spin 1/2 nuclei ^{13}C is of interest in chemistry, especially organic chemistry. However the isotope ^{13}C has a natural abundance of only 1.11 % as compared, for example to 99.98 % for ^{1}H. ^{12}C has an abundance of greater than 90 % but spin $I = 0$. In spite of its low abundance ^{13}C NMR is frequently used in chemistry. ^{13}C chemical shifts are in the range 0-230 ppm with reference to TMS. Many of the factors that affect the ^{13}C chemical shifts are the same as for ^{1}H NMR. The state of hybridisation of the observed ^{13}C nucleus is an important factor. sp^3 carbons resonate between -20 and 100 ppm with respect to TMS; sp^2 carbons between 120 and 240 ppm and *sp* carbons 70 and 110 ppm. The electronegativity of the substituent is another important factor influencing the chemical shifts, as are steric and van der Waals effects. Mesomeric effects arise from electron-donating groups attached to benzene rings that shield the ortho and para carbon atoms (electron withdrawing groups would deshield these positions). Hydrogen bonding, attachment of heavy atoms, bond conjugation, attachment of deuterium atoms instead of hydrogen, *pH* and the solvent, all contribute to the ^{13}C chemical shift.

Spin-spin coupling in ^{13}C NMR can be observed between the ^{13}C nucleus and neighbouring protons, as well as between two ^{13}C nuclei. The latter effect is more difficult to observe, since due to the low natural abundance of ^{13}C isotopes, the chances of two such nuclei occuring as neighbours in the molecule is very small.

^{13}C NMR also lends itself to the study of sample in the solid state. In solutions, local magnetic fields due to the nuclei of neighbouring molecules will be averaged out due to the rapid tumbling motion and will not have any effect on the width of the resonance peak. In solids however, each individual molecule would be surrounded by a different static environment. This leads to a widening of the resonance. Until the advent of recent high resolution solid state NMR methods, ^{13}C NMR was only sparsely applied in the solid state, and many interesting aspects of solid state chemistry, polymer chemistry, etc. could not be studied. High-power decoupling cross polarisation (CP) and Magic Angle spinning (MAS) are two of the relatively recently developed techniques which make it possible to obtain high resolution even in solids. So much so, at least for low molecular weight compounds, solid state NMR gives a resolution comparable to the solution state.

8.8 NMR Application in Biochemistry and Biophysics

In addition to the fact that biologically important molecules are usually of very large molecular

weight and size, biological systems also reveal more complex interactions between the molecules or groups of molecules. The study of such complex systems through NMR has in recent times become possible in much greater detail and precision with the advent of very high-field magnets and techniques such as multi-dimensional NMR. Even before these developments, however, NMR was being used in biochemistry to determine concentrations, study the pH behaviour of biomolecules and to elucidate the conformation of small biologically important molecules.

8.8.1 Concentration Measurement

The determination of relative concentration by NMR is made simply by measuring the area under the NMR absorption signal. If the spectrum has been calibrated against a signal of known concentration, absolute concentrations can be measured in this way. Precautions however have to be taken against several sources of error that may creep in. One of these is the fact that if, due to different relaxation times, one signal has become saturated, while another is not, the measurements of the relative concentrations may be wrong. Another source of error is the possible uneven intensity of the excitation pulse which may lead to different frequencies being excited with different intensities. It is however possible to choose signals which do not suffer from such drawbacks, enabling fairly precise measurements of concentrations.

8.8.2 pH titration

The protonation or deprotonation of functional groups in biological molecules has a detectable influence on the chemical shift. The rates of proton exchange with the solution are much faster than the NMR time scales and the average effect is observed. This effect is not only sufficiently strong to be detected at the protonation/deprotonation site, but even several bonds away in the molecule. The protonation of a particular site depends on its *pK*. pH titration is the measurement of the chemical shift as a function of pH and helps to measure the *pK* of the site. The protonation or deprotonation of the site at its *pK* value is marked by a sharp change in its chemical shift at that value of the pH. In proteins, histidine residue resonances are especially dependent on the pH of the solution and together with other such resonances can be used as a monitor of denaturation. This is because protonation sites in the interior of the proteins will be progressively exposed to the solution as denaturation proceeds, which can be detected by the dependence of the spectrum on pH or temperature (or other denaturing conditions). Inorganic phosphates participate in many enzymatic reactions, and exist, during the course of the reaction, in many different protonated states such as H_3PO_4, HPO_4^-, HPO_4^{2-}, and PO_4^{3-}. By using ^{31}P NMR and pH titration, the course of such a reaction can be followed.

8.8.3 Conformation of Biomolecules

The NMR spectra of biomolecules contain a large amount of structural and conformational information. The problem is to extract this information. Even at relatively high resolution, most of the peaks in a ^1H NMR spectrum of a protein are broad. The causes for such broadening in macromolecules are as follows. (a) A biopolymer usually contains more than one unit of any particular monomer, and each unit would be in a slightly different chemical environment, leading to a widening of the resonance. This cause for the widening is the same as that occurring in solids. (b) The presence of other nuclei close-by leads to dipolar relaxation in excess of what would be observed for the monomer and contributes to the broadening of the resonance. (c) Chemical exchange of protons between different close-by sites can also cause line broadening. (d) Paramagnetic ions are frequently found in biological molecules and are a source of line broadening. The consequence of all this is that detailed conformational

studies are possible by the conventional one dimensional and or continuous wave NMR only in the case of small organic molecules (as discussed earlier). However, certain aspects of biomolecular conformation can be studied by conventional NMR. One such is the helix to random-coil transition in polypeptide chains. Similarly one can use NMR to follow the denaturation-renaturation transition in proteins. In either case the NMR spectrum of the random coil would be a composite of the spectra of the constituent amino acids. The spectrum of the ordered structure would deviate markedly from the simple random-coil spectrum. The transition can thus be followed as function of temperature, *pH* or any other relevant parameter.

8.8.4 Two-dimension NMR

As mentioned earlier, detailed conformational studies of macromoleules are possible now with the development of two-dimensional techniques. A one-dimensional FTNMR experiment involves the application of a RF pulse after the sample has been allowed to achieve equilibrium during the 'preparation' period (Figure 8.10(a)). Immediately after comes the detection phase where the FID (i.e. the signal S is collected as a function of time (here represented by the variable t_2). A Fourier transform of the signal into the frequency domain results in the NMR spectrum.

$$S(t_2) \rightarrow S(\omega_2)$$

In the case of two dimensional NMR spectroscopy an additional time period (denoted t_1) and called

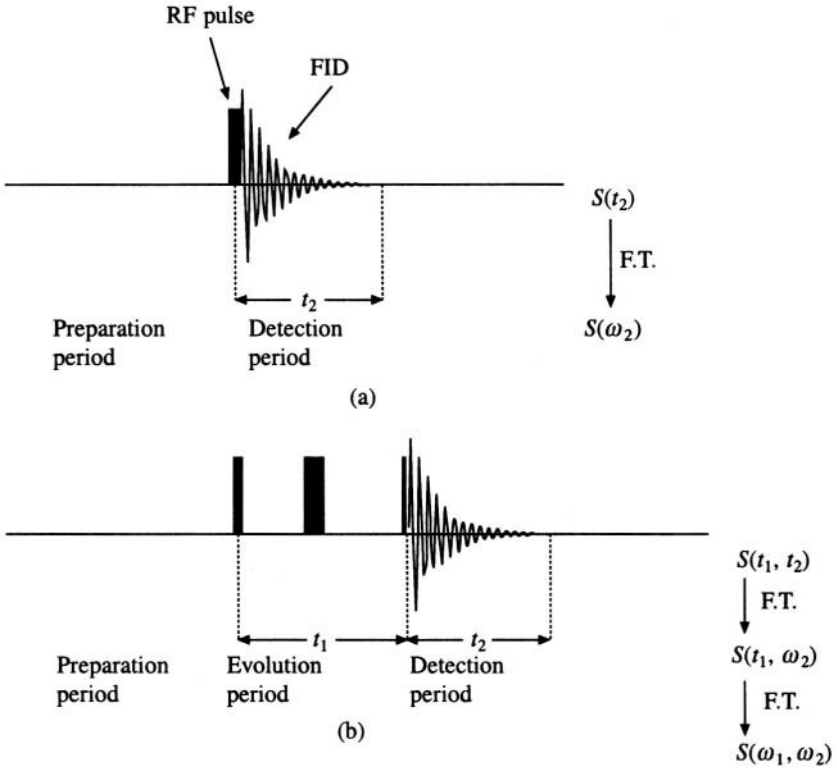


Fig. 8.10. A schematic representation of the NMR experiment. (a) 1-D experiment. (b) 2-D experiment.

the evolution period is introduced, after which the FID is collected (Figure 8.10(b)). During the evolution period another RF pulse may be applied and a further period of time, called the mixing period, may be allowed to elapse before the FID is collected. In each experiment, the time t_1 may be given a different value, ranging from zero to some maximum value depending on the relaxation times. Each time the FID is collected and Fourier transformed to yield $S(t_1, \omega_2)$. The several spectra so obtained are arranged in a data matrix which can be subjected to a second Fourier transform, column wise

$$S(t_1, \omega_2) \rightarrow S(\omega_1, \omega_2)$$

The spectrum thus will have two dimensions. ω_1 is along one of the dimensions and ω_2 along the other. The ω_2 axis contains information about coupling interactions with other nuclei during the evolution period t_1, and the ω_2 axis contains information about the chemical shifts. The precise pulse sequence applied to the samples, gives rise to different types of coupling information being displayed along the ω_1 axis. There are currently more than 50 different pulse sequences that have been published. The two most commonly used sequences in biological research are abbreviated as COSY and NOESY.

COSY is an abbreviation for correlated spectroscopy. There are different types of COSY pulse sequence. The so-called homonuclear shift correlated spectrum displays the *J* coupling, or the spin-spin coupling through bonds as a two-dimensional plot. Figure 8.11 is a simple COSY spectrum in which the connectivity of spins in the molecule can be mapped out. COSY spectra are of great use in assigning the peaks to the nuclei that cause them. Solving the chemical structure of the molecule can be accomplished through a COSY spectrum. The assignment of peaks in the spectrum of a large biomolecule such as a protein is relatively easily accomplished.

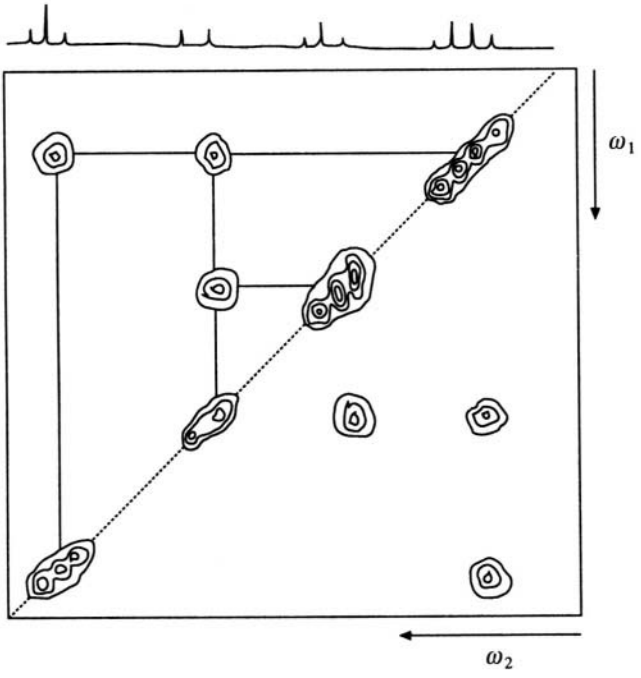


Fig. 8.11. A simple COSY contour plot shows the connected atoms.

NOESY is an abbreviation for Nuclear Overhauser Enhancement spectroscopy. The nuclear Overhauser effect is a 'through-space' effect and can be used to determine the spatial separation between nuclear spins. NOESY spectra are similar to COSY spectra except that in this case the off diagonal peaks correspond to 'through-space' connectivity of spins. The strength of the effect varies as $1/r^6$ where r is the distance between the spins, and measurement of the strength of the off diagonal NOE peaks in a 2D NOSEY spectrum can be used to estimate distances between spins. As explained in the next section, such measurements are of great use in arriving at the three-dimensional folding patterns of biopolymers such as proteins in which the polypeptides chain would fold back on itself to bring into close proximity nuclei which are widely separated along the chain.

8.8.5 Determination of macromolecular structure

The experiment to determine the three dimensional structure of a macromolecule usually consists of two steps. The first step is to assign all the resonance peaks that occur in the spectrum. The high-resolution spectrum of a macromolecule such as a protein is very complex and only through a 2-D spectrum such as COSY is it possible to trace the connectivity in the molecule and assign the peaks. In proteins, especially, but also in nucleic acids, the interpretation of the 2-D COSY spectrum is sometimes made easier by the existence of recognisable patterns of connectivity which can be attributed to specific amino acids or base-sequences. The second step in the structure determination is to identify secondary and tertiary structures from the *J*-connectivity patterns and from NOE data. A NOESY spectrum will give information about the 'through-space' couplings between the spins in the molecule. The NOE intensities may be divided into classes as strong, medium and weak NOEs corresponding to a distance less than 2.5 Å, 3.0 Å and between 3.0 and 4.0 Å, respectively. (Usually no NOEs are observable beyond 4.0 Å). The distance data so obtained may be used to attempt a reconstruction of the three-dimensional folding of the biopolymer chain. This is mathematically a complicated problem and no general algorithm exists which can lead from the distances to the geometry in a reliable way. There are however several approaches to solving this problem, each being successful to a lesser or greater degree. The NOE distances alone are not sufficient to define the structure. Oxygen and nitrogen atoms, for example, are usually not seen in the NMR spectra and their positions must be inferred from the positions of connected protons or ^{13}C nuclei. Distance geometry algorithms therefore use all chemical information available and may be performed in torsional-angle space where the degrees of freedom are the dihedral angles of the main chain and the side chains. Additional constraints in the form of non-bonded van der Waals contacts, hydrophobic interactions and so on, may also be used in arriving at the structure. A commonly used technique is restrained molecular dynamics. As described elsewhere in the book the observed experimental data, in this case the NOE distances, are written as a pseudo potential, along with other semi-empirical potential terms such as bond length energy, bond angle energy etc. This method has been successfully used in many instances in arriving at precise three-dimensional structures of small proteins (Mw ~20,000). Structures of larger proteins (Mw > 50,000) are still somewhat inaccessible to current techniques.

8.9 NMR in Medicine

Some of the most spectacular applications of modern NMR techniques have been in the area of medical imaging, namely Magnetic Resonance Imaging or MRI. Unlike NMR spectroscopy, where the NMR signals are obtained as a single spectrum from the entire sample, imaging requires that the signal be resolved in space. The different signals from the different parts of the object are reconstructed on the computer screen as an image of the object. The electromagnetic fields and waves used in NMR pass easily through living bodies and signals from any interior portion can be detected and processed.

In this way it is possible to build up clear images which will be of great use in medicine. Three NMR measurable parameters are commonly used in imaging. The first is the spin-density or the number of a particular type of nucleus (usually the water protons) per unit volume. This parameter can be deduced from the intensity of the corresponding resonance peaks at all the different portions of the object. The second parameter that can be used in imaging is the spin-lattice relaxation time T_1. Different types of tissue within the object would have different T_1 values and if these values are mapped out, an image is obtained. The third commonly used parameter; viz. the spin-spin relaxation time T_2 is also treated the same way. Thus, in contrast to X-ray imaging (including the CAT scan), where the X-ray absorption is the only detectable parameter, there are at least three parameters which can be detected in NMR imaging and NMR tomography. This allows greater detail and also different types of tissues to be imaged. In addition, NMR does not use any highly ionising radiation to produce its images.

The most basic principle in the imaging technique is the selection of a slice to be imaged from the entire volume of the object. Unlike as in NMR spectroscopy, where the magnetic field is kept steady over the entire sample, in the case of NMR imaging, spatial selection is achieved by applying a magnetic field gradient on top of the static magnetic field. This would mean that particular spins, such as for example protons (^{1}H) come into resonance at different frequencies at different parts of the sample along the axis of the field gradient. If the sample is irradiated by a radio-frequency excitation pulse in a very narrow range of frequencies, then only the spins in a particular spatial slice of the object, perpendicular to the direction of the magnetic field gradient would come into resonance and contribute to the NMR signal. If a 2D image of such a slice is obtained, many such images can be stacked one on top of the other to yield a complete image of the object.

In order to obtain a 2-dimensional image of the slice, a 2D-Fourier imaging technique is most commonly used. The slice selection gradient is assumed to be applied along the z axis. Thus the slices are perpendicular to this axis and parallel to the xy plane (Figure 8.12). After the first excitation pulse, a gradient $G(x)$ is applied for a time t_x along the x-axis. Immediately after, another gradient $G(y)$ is applied along the y-axis for a time t_y. The NMR signal is then collected. The process is repeated for different values of t_y ranging from 0 to the maximum, each time obtaining the signal for different values of t_y. Thus a matrix of signal of size n^2 is obtained where n is the number of different values of t_x and t_y. It can be shown that the matrix when Fourier transformed will result in 2D image of the slice. The time taken to obtain such an image is usually less than 10 minutes. While this is sufficiently short for imaging live animals, if the process were to be repeated several times in order to obtain the complete three dimensional image, it is difficult to keep the object still and motionless for the few hours that it might take. Other methods have been developed where additional slices are measured almost simultaneously and the entire image reconstructed in one go. This may entail a loss in resolution, contrast, etc.

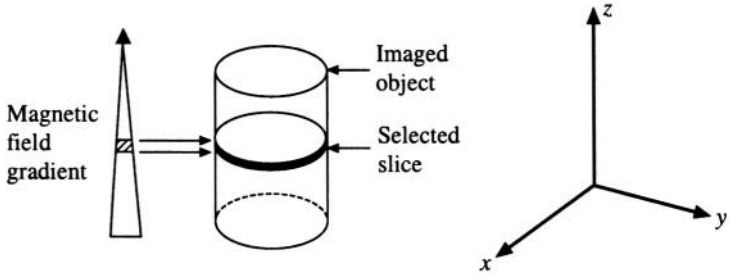


Fig. 8.12 The geometry of the MRI experiment.

One of the most important applications of NMR tomography is the imaging of sensitive parts of the human body such as the head. X-rays are highly attenuated by the skull, making X-ray imaging of the brain quite difficult, in addition to the fact that X-rays are ionising radiation and therefore can cause damage. The radio frequency radiation used in NMR passes almost unhindered through the skull and has no as yet known ill effects on the tissues. Certain images of the skull, such as the so-called sagittal view are simply not possible with X-ray tomography. In addition, pathological conditions are easily identifiable especially with the possibility of using different modes of imaging (i.e. spin density, T_1 or T_2). Apart from these three parameters it is also possible to image the organs by means of the difference in the chemical shifts. NMR imaging may also be used to study continuous flow processes within the body such as blood flow. The excitation pulse sequences may be adjusted so that only the flow is visible, leading to a clear image of the blood vessels alone.

NMR imaging at very high spatial resolution is called NMR microscopy. At present the maximum obtainable resolution is about 20 μM, mainly because below this limit there are simply not enough protons in the sample volume to give a signal with sufficiently signal-to-noise ratios. Thus NMR microscopy, while potentially of great use, since the sample is preserved intact even while observing interior portions, has yet to reach the resolution of the best light microscopes.

A related application of NMR in medicine is magnetic resonance spectroscopy (MRS) also known as 'in vivo spectroscopy'. Once again a spatial selection technique is used to obtain a reasonably high resolution NMR spectrum from a small volume of the object, such as typically, the human body. MRS differs from MRI in that in the latter, information is available regarding only one aspect of the sample such as the spin density. MRS looks at the entire information content of the spectrum. The spatial selection techniques are somewhat different in detail from those used in imaging. This is because the portion of the sample volume may be defined with respect to either the laboratory frame of reference or with respect to certain biological structures, i.e. whether the molecule investigated is inside or outside a defined closed structure such as a cell. Typically in vivo spectroscopy is used to monitor the state of various identifiable metabolites as a function of the progress of some biological condition, e.g. disease. This method is thus in the process of developing into a sophisticated early diagnostic tool as well as a way of following the results of therapy. The method however is only now moving out of the laboratory and into the clinics.

9

Molecular Modelling

9.1 Introduction

X-ray analysis or NMR or other methods of structural characterisation of the molecules give, as output, a set of numbers that specify the relative positions of each of the atoms of the molecules in space. In order to use the information present in that list, human capabilities almost always demand that it be converted to a form in which it can easily be visualised. Before the advent of powerful computer graphics, the co-ordinates were used to build models in wood or plastic. These suffered from the following disadvantages: (1) They were tedious to build. (2) They were static; they could not be easily altered to try out new ideas about the structure. (3) They were not accurate. (4) They were fragile. (5) They were heavy, prone to collect dust and were difficult to maintain. (6) They took up a lot of space. (7) They were expensive.

Despite these disadvantages they were widely used and several new ideas in biology can be credited to such models, the most notable being the double-helical structural model of DNA, constructed by J.D. Watson and F.H.C. Crick with metal parts.

Computers have made all the difference to molecular modelling. They possess none of the disadvantages listed above. The chief problem with computer graphics is that they are two-dimensional pictures of 3-dimensional objects. Nevertheless one can get over this problem to a great extent by using techniques such as stereopsis. Most programs also have a facility which allows the user to 'rotate' the picture on the display screen so that different parts of the molecule come into view continuously giving the illusion of watching a real three-dimensional model. Thus the advantages of using computer graphics far, far, outweigh the single disadvantage.

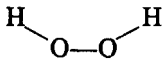
9.2 Generating the Model

If the co-ordinates of the molecule are already available, from, say, an X-ray crystallographic analysis, then they are just fed into the computer program which will interpret each data point as an atom and

display it according to the mode and orientation chosen by the user. If the co-ordinates are not available, but if the chemistry of the molecule is known (i.e. the chemical formula for a small molecule; and the amino acid sequence for a protein), it is possible to generate the co-ordinates for the molecule, *ab initio,* on the computer. As an example we demonstrate the generation of the co-ordinates for a very simple molecule, namely, hydrogen peroxide

9.2.7 *Building the structure of* H_2O_2

Hydrogen peroxide, H_2O_2 has the chemical structure



Chemical studies have established that the molecular dimensions are as given below:

Distances	Angles
O–H = 1.1 Å	H–O–O = 120°
O–O = 1.5 Å	

The one variable in this molecule, which is left to the choice of the modeller, is the dihedral angle[8] or the torsion angle H–O–O–H. If this is changed, it implies a rotation of one hydrogen atom about the O–O bond, with the other hydrgen atom kept fixed. In order to generate the co-ordinates for such a molecule, we first define the frame of reference as in Figure 9.1. Note that it is conventional to

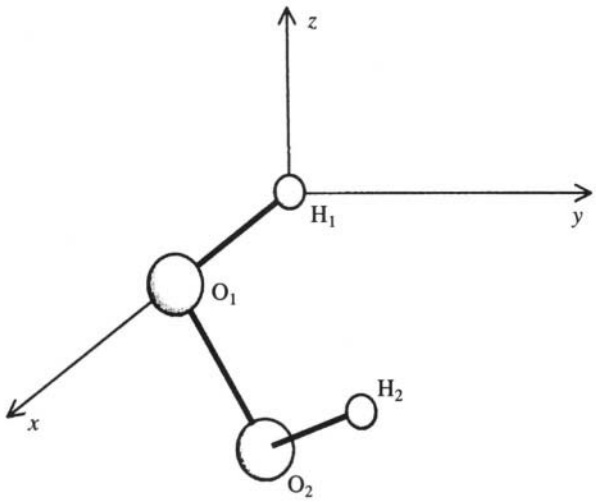


Fig. 9.1. Frame of reference chosen to generate H_2O_2

[8]A dihedral angle is the angle between two planes, measured as the angle between the normal vectors to the two planes. A torsion angle is the angle made between the two outermost bonds when three consecutive bonds are considered. For example, consider the atoms *P, Q, R* and *S* bonded in a chain, *P* to *Q, Q* to *R* and *R* to *S.* The torsion angle between *P–Q* and *R–S* is measured by first viewing the four atoms along the central *Q–R* bond. In this view, atoms *Q* and *R* superpose on one another. The angle made by the far bond, say *R–S,* with respect to the near bond, *P–Q,* measured in the clockwise direction, is the torsion angle. This is the same as the angle between the plane *PQR* and the plane *QRS,* and hence is sometimes called the dihedral angle.

choose a right-handed Cartesian co-ordinate system. We first place one of the hydrogen atoms, which we call H_1, at the origin. Co-ordinates of H_1 are therefore

H1 0.0 0.0 0.0

Next we place the O1 atom along the x-axis, at a distance of 1.1 Å from orign. The co-ordinates of O1 are therefore

O_1 1.1 0.0 0.0

The third atom O_2 placed in the *xy* plane, at a distance of 1.5 Å from O1 and such that the O_2–O_1 bond makes 120° with the H_1–O_1 bond. An elementary geometrical analysis shows that the co-ordinates of O_2 are

O_2 1.5 1.3 0.0

Finally in order to determine H_2, we place it at a O_2–H_2 distance of 1.0 Å and O_1–O_2–H_2 angle of 120°. This still leaves us the freedom to place the atom anywhere on a circle. If we choose the torsion angle to be + 60, the co-ordinates of H_2 are

H_2 1.7 2.0 0.8

Thus the spatial positions (co-ordinates) of all the atoms in hydrogen peroxide molecule can be obtained by simple geometrical considerations. The lengths between the bonded atoms, the angles and the torsion angles, are called the internal parameters or internal co-ordinates of the molecule. A complete set of internal co-ordinates is exactly equivalent to the set of Cartesian co-ordinates and one can be derived from the other in the complete and non-redundant way.

9.2.2 Building protein structure

The geometry of each amino acid in the protein is known, i.e. the relevant bond lengths and angles are known. Also known is the geometry of the linking peptide bond. Despite this, building a protein chain is not a trivial task since the torsion angles are not known. In the amino acids with the longer side chains, especially, the orientations of the various chemical groups depend on the side chain torsion angles. Apart from this, in the main chain, or the peptide backbone, the ϕ and the ψ angles are not fixed by chemistry. Hence, unless the fold of the protein is known or guessed at, a three-dimensional model of the entire protein molecule cannot be built.

It is known that information regarding the final three-dimensional structure of the protein is inherent in its amino acid sequence. Hence knowledge of the latter should, in principle, lead us to the former. However, the rules of protein folding are still unknown. It is not yet possible to go from the sequence to the structure without additional experiemental information. This is called the 'Protein Folding Problem' and is one of the great, unsolved problems in molecular biology. It may be succinctly stated as follows: Given the amino acid sequence of a protein, how does one predict its three-dimensional structure. Several methods of structure prediction are available, though none of them have been completely successful. One of the most popular prediction methods was developed by P.Y. Chou and G.D. Fasman. The Chou and Fasman method has been quite successful in predicting the secondary structure of proteins. The method is based on a statistical analysis of the known protein structures that have been collected in a computer databse.

Extensive collections of published three dimensional structures are now available in computer readable format. A collection of all the published protein structures is available at the Research Collaboratory for Structural Biology in USA and is called the Protein Database or PDB for short. A

typical entry in this databse would consist of the name of protein, the authors of its structure, the methods used to determine the structure, and a list of the coordinates and thermal parameters of each atom in the structure including solvent atoms, ligands, etc. There are now over fifteen thousand entries in the database. Most of the structures deposited in the database have been solved by X-ray crystallography. There are also about 500 structures determined by NMR techniques. The number of structures in the database is increasing rapidly and about 2000 new structures are added to the database every year. The database can be accessed through computer networks or can be obtained on CD-ROMS. A similar high quality database is the crystallographic structural database or the CSD at the University of Cambridge, UK. This contains more than 1,00,000 entries of the co-ordinates of the crystal structures of small organic molecules, i.e. those with a molecular weight less than about 1000.

Chou and Fasman used the protein structural database to estimate the frequency with which each amino acid occurred in one of the three main secondary structural motifs, i.e. **α-helix, β-strand** and reverse-turn. The normalised values that they obtained are given in Table 9.1.

Table 9.1 Chou-Fasman propensity values of the 20 amino acid residues for each of three secondary structures

Residue	α-helix	β-strand	Reverse turn
Ala	1.45	0.97	0.15
Cys	0.77	1.30	0.31
Asp	0.98	0.80	0.33
Gln	1.53	0.26	0.12
Phe	1.12	1.28	0.19
Gly	0.53	0.81	0.45
His	1.24	0.71	0.18
Ile	1.00	1.60	0.1[illegible]
Lys	1.07	0.74	0.27
Leu	1.37	1.22	0.14
Met	1.20	1.67	0.18
Asn	0.73	0.65	0.45
Pro	0.59	0.62	0.41
Glu	1.17	1.23	0.15
Arg	0.79	0.90	0.27
Ser	0.79	0.72	0.41
Thr	0.82	1.20	0.27
Val	1.14	1.65	0.08
Trp	1.14	1.19	0.30
Tyr	0.61	1.29	0.33

The numbers represent the propensity for each amino acid to take up the respective conformation. Thus glutamine has a very high propensity to occur in a helix, i.e. to be a helix-former. Similarly isoleucine strongly prefers to exist in a **β-strand.** The propensities for the reverse turn are not as well expressed. However, it may be noticed that some residues like glycine or asparagine have a fair degree of propensity for the reverse turn, without a strong preference either for the **α-helical** or **β-strand** conformation. The way this table is used to predict the secondary structure of a polypeptide chain is illustrated by means of an example.

Consider a 18-residue sequence such as shown in Figure 9.2. The **α-helical, β-strand** and reverse turn propensities for each residue is also indicated as taken from Table 9.1. Also indicated in the

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
Sequence	Lys	Ile	Thr	Val	Val	Gly	Ala	Gly	Ala	Asn	Gly	Leu	Ala	Gln	Ala	His	Ser	Leu
Propensities																		
α-helix	1.07	1.00	0.82	1.14	1.14	0.53	1.45	0.53	1.45	0.73	0.53	1.37	1.45	1.53	1.45	1.24	0.79	1.37
β-strand	0.74	1.60	1.20	1.65	1.65	0.81	0.97	0.81	0.97	0.65	0.81	1.22	0.97	0.26	0.97	0.71	0.72	1.22
β-turn	0.27	0.15	0.27	0.08	0.08	0.45	0.15	0.45	0.15	0.45	0.45	0.14	0.15	0.12	0.15	0.18	0.41	0.14
Average Propensity (window = 6)																		
α-helix	1.01	1.03	0.95	0.96	0.91	0.97	0.92	0.82	0.96	1.01	1.18	1.18	1.26	1.31	1.31	1.28	1.21	1.13
β-strand	1.30	1.37	1.28	1.43	1.30	1.26	1.09	0.95	1.02	0.91	0.81	0.81	0.82	0.81	0.81	0.78	0.91	0.88
β-turn	0.19	0.17	0.22	0.19	0.24	0.22	0.28	0.34	0.29	0.30	0.24	0.24	0.20	0.19	0.19	0.20	0.22	0.24
	|←———	β-strand	———→				|←———	β-turn	———→		|←———			α-helix	———		———→	|

Fig. 9.2 Secondary structure prediction of a 18-mer oligopeptide using the Chou-Fasman method

figure are the average values of the propensities taken six at a time. (Hexa-peptides are the normally used probe length, though other lengths also may be used. For the **α-strand** especially the probe length can be as short as five peptides). An examination of the averages shows that the average strand propensity is the highest for the hexa-peptide that starts at residue 1. A **β-strand** can be considered to nucleate at this point, since the average propensity for an **α-helix** is rather low at the same point. The strand can be extended in both directions until a region is reached where the helix is a more probable structure. Thus on comparing the propensities for both types of structure it can be determined that the residues 1 to 6 are most likely to be **β-strand,** while residues 11 to 18 are most likely to be **α-helical.** The determination of possible reverse turns is more complicated since each residue has a propensity value for its occurrence at each position of the reverse turn. (A reverse turn has four residues i.e. four positions). The principle however is the same, and, after a thorough analysis, the entire polypeptide chain can be classified as **α-helix, β-strand,** reverse turn or none of the above, i.e. random coil. Such methods of secondary structure prediction are now available as easy-to-use computer programs. Given the sequence of a protein one can use the program to quickly predict its secondary structure.

The problem of arriving at the structure of the protein from knowledge of its sequence is nevertheless far from solved. Firstly the secondary structure prediction methods, such as the Chou and Fasman method are still quite approximate, and all the regular structural elements of the protein chain cannot be identified. Predictions made have often turned out to be wrong and the overall rate of success does not exceed 60% in the best test cases. Methods are still evolving and combinations of methods are expected to have much higher success rates.

The other major obstacle in protein structure prediction is that the arrangement of secondary structural elements in space follows rules that are not completely understood. This comes in the way of arriving at the tertiary fold of the chain from a knowledge only of the sequence. The prediction of the three dimensional structure of a series of proteins with homologous sequences is much easier if the structure of one of them is known. Thus almost the first task that is done when the sequence of a protein is known to compare it with all other known protein sequences and identify those which are similar. If the structure of one of the similar proteins is known through other experimental or theoretical techniques, it can be used as a guide in generating the structure of the new protein. The principle which is followed is that stretches of similar sequences must have similar or the same secondary structure. If the two sequences show a high degree of overall similarity (homology), then the tertiary fold of the two sequences can also be taken to be similar, and the co-ordinates of the unknown protein can be generated by appropriately modifying the co-ordinates of the known protein. This method of generating protein structures is especially used to build the initial trial model in the molecular replacement method in X-ray crystal structure analysis.

9.2.3 Building nucleic acid structure

To build a DNA double helix is relatively simple as compared to building a protein chain. All one needs are the co-ordinates of one nucleotide unit and the helical transform that generates the next repeat along the strand. A rotation about the two-fold axis perpendicular to the helix axis generates the other strand. Monomer co-ordinates for A, B and Z type DNA are available in the literature. Other unusual, regular DNA structures such as tetraplexes, triplexes, etc., can also be generated by following the same principle. "Real" DNA molecules are not perfectly regular and show a sequence dependent variation from the standard structure at each base-pair step. Since no data on the correlation between the sequence and the microstructure exists, it is difficult to build models that show this variation. Thus most DNA models are regular helices.

Models of RNA structures are much more difficult to build. Only a few structures of large single

stranded RNA molecules exist and the principles of RNA structure are not well known. An analysis of the sequence of the RNA molecule, if known, enables prediction of secondary structure, specifically regions of the molecule which base pair with each other and therefore can be thought to exist as short double-helical stretches. The way in which these double helices are put together to form the three-dimensional RNA structure is an unsolved problem.

9.2.4 Displaying and altering the generated model

There are many ways of representing the three dimensional molecule on the two dimensional computer display. Choosing the appropriate mode of representation depends on the particular information that is required to be emphasised.

Small molecules are typically represented by the so-called ball and stick model (Figure 9.3), where each atom is shown as a sphere and each covalent bond as a stick. The radius of the sphere is commonly taken to be proportional to the Van der Waals radius of the respective atom. In colour representations, a commonly used code is hydrogen—white, carbon—black, nitrogen—blue, oxygen—red, phosphorous and sulphur—yellow. If the anisotropic thermal vibrations of the atom also need to be represented, then ellipsoids are used instead of spheres to depict the atoms (Figure 9.4). The radius of the atoms may be increased to an extent where the sticks representing the bonds are hidden. Such a model is called a space-filling representation. A chemically accurate version of a space-filling model was first designed in plastic by Corey, Pauling and Koltun and is called the CPK model and may be thought to represent the electron could of the molecule. A computerised version of this draws the CPK model on the screen. Large molecules, such as most proteins, are effectively seen when drawn as line diagrams rather than ball and stick models. Even then, there is too much detail to be fully appreciated (Figure 9.5). One way of getting over this, problem may be to show the picture in stereo (Figure 9.6). It may be sufficient to show only the atoms of backbone, if the interest is only in

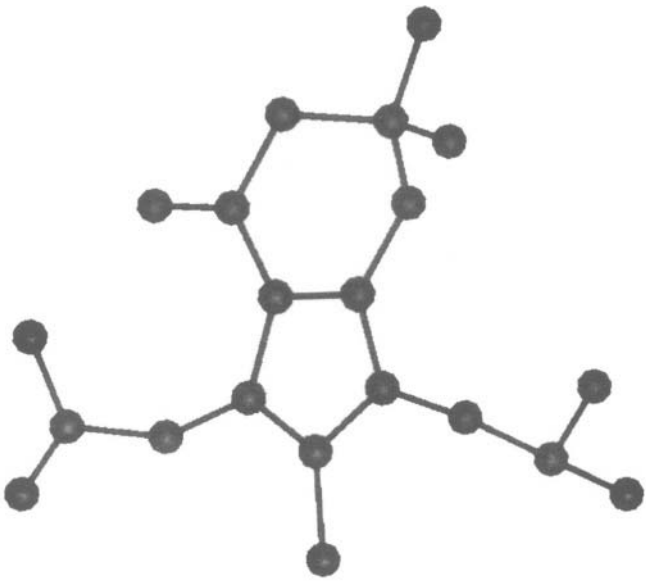


Fig. 9.3. A ball-and-stick representation of an indole molecule showing all the atoms by spheres of equal radius.

the protein fold. A further abstraction is to plot and join only the C^{α} atoms (Figure 9.7). Cartoon diagrams (Figure 9.8) are useful to show the disposition of the seondary structural elements. It is conventional to depict **α-helices** as cylindrical rods and **β-strands** as flat arrows. If the interest is only in the surface of the protein, diagrams such as Figure 9.9 may be used.

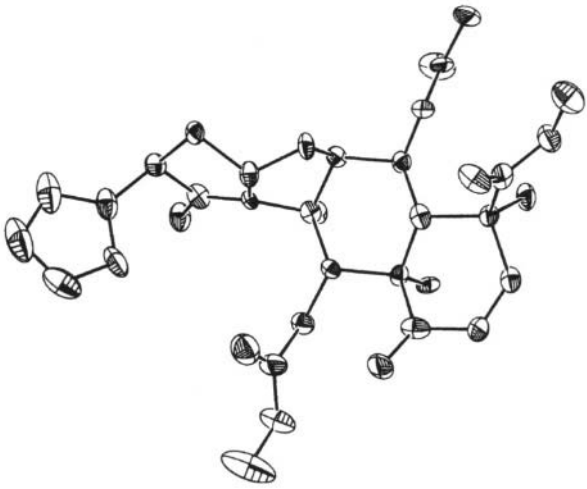


Fig. 9.4 A thermal ellipsoid representation of a molecule of thermal vibration of each atom, indicating the extent (drawn using the computer program ORTEP).

9.3 Optimising the Model

The generated molecule needs to be checked for compliance with the laws of physics and chemistry before it can be accepted for further processing. The stereo-chemical criteria used during the construction of the model are usually insufficient to decide upon the final correct structure. The dynamics of the molecule, i.e. the various forces acting on various parts of it have to be explicitly calculated and the structure has to be modified to optimise the action of these forces. A simple first step in the structure optimisation is to ensure that the bond lengths and bond angles fall within the ranges measured in accurate spectroscopic or X-ray diffraction experiments. Atomic sizes are also known from experiment and the fold of the generated molecule should not be such that the distance between two non-bonded atoms is less than the sum of the atomic radii, i.e. there should be no inter-penetration of the atomic spheres. These chemical and physical data can be checked and optimised using computer programs. It is usually not possible to exactly match the requirements. However the difference between the expected and observed geometrical values is minimised using least squares techniques.

Such geometrical optimisation is sufficient to give a plausible model. It is however not sufficient to serve as a basis for new interpretation regarding the structure-function relationships in the molecule, i.e., no new 'science' can be expected from it. Energy minimisation and structural optimisation using molecular dynamics are at the next higher level of modelling. The free energy of the molecule is an experimentally measurable thermodynamic quantity. This quantity can also be computed. A very important physical principle requires that at the equilibrium state of any system, the free energy be the least. In other words, the minimum energy state is the equilibrium state and all systems tend to move towards this state. This process of structure optimisation is used to determine the minimum energy configuration. In order to calculate the free energy for various states or various conformations

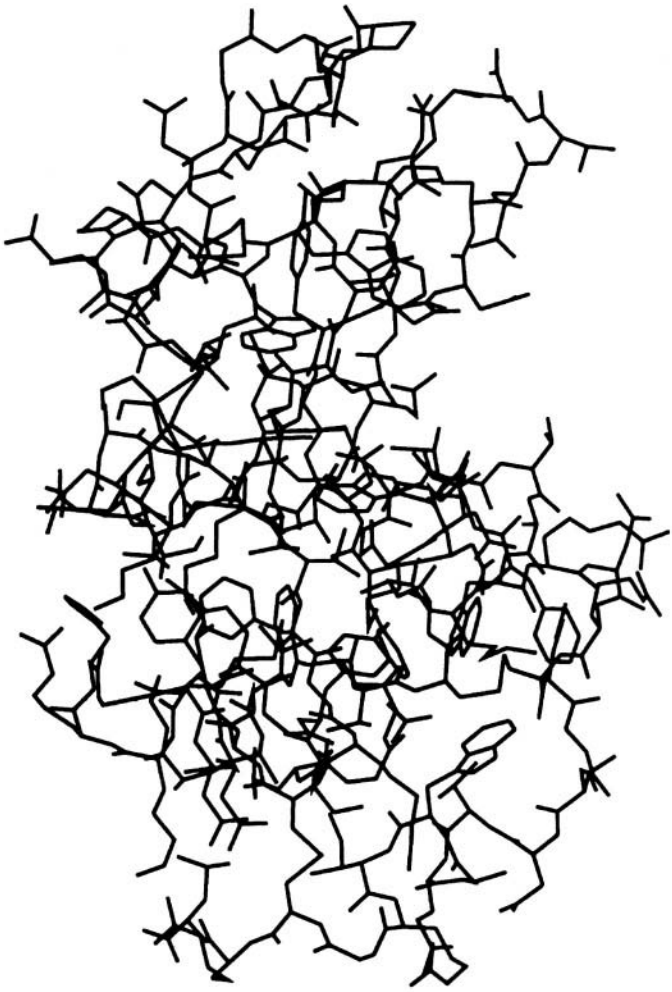


Fig. 9.5. Line diagram of lysozyme

of the molecule it is necessary to set up the quantum mechanical equations for the whole molecule and to solve them. This task is beyond the scope of currently available mathematical and computational techniques. However, the value of the free energy can also be calculated from a semi-empirical expression such as

$$E_{tot} = E_{bond\text{-}length} + E_{bond\text{-}angle} + E_{tor} + E_{non\text{-}bonded} + E_{elec} + E_{hydrogen\text{-}bond}$$

where E_{tot} is the total semi-empirical free energy, $E_{bond\text{-}length}$ is the contribution to the energy due to the deviation of the bond lengths from their expected values, $E_{bond\text{-}angle}$ is the contribution due to the deviation of the angles from their standard values, E_{tor} is the contribution due to the torsion angles and is dependent on the disposition of the four atoms that make up the torsion angle (e.g. an eclipsed configuration has a very high energy). $E_{non\text{-}bonded}$ is the contribution due to the van der Waals contacts between the atoms. Interpenetration will raise the energy enormously. E_{elec} is the contribution due to the electrostatic interaction of attraction and repulsion between the atoms. Even though the molecule

as a whole may be electrically neutral, quantum mechanics tells us that each atom will posses a 'partial' charge, positive or negative, and it is these partial charges which take part in the electrostatic interactions. This is in addition to interactions between charged residues such as Glu or Asp. Finally $E_{\text{hydrogen-bond}}$ is the energy contribution due to hydrogen bonds. Other terms, such as one for the 'anomeric effect' are sometimes added to this expression for the energy. The conformation of the molecule is optimised with respect to the energy. In other words, the conformation is changed so that the value of the energy is a minimum. The minimisation is carried out by some well-known mathematical techniques such as least squares or gradient search methods. Such a procedure will give the final 'correct' conformation of the molecule. In addition new interactions and structures not built into the original model may appear.

Fig. 9.6. Stereo picture of protein. (Though only a two-dimensional image of the object falls on each retina, the brain is able to synthesize the three-dimensional picture from the two images since they are produced by views at slightly different angles. Stereo pictures mimic this effect by providing views of the molecule as seen from two different angles. In order to obtain the 3-D effect it is necessary to allow the two eyes to simultaneously focus only on one picture each, to the exclusion of the other. This may be done by placing a card between the eyes and blocking out the second image. Many people are able to obtain the 3-D effect without any such aids. The reader may try to stare at the example given above, focusing and defocusing his/her eyes unitl he/she obtains the 3-D effect. The first attempt may take hours before it is successful.)

The molecule may also be subjected to the molecular dynamics calculations to study its dynamical behaviour. In this procedure each atom is thought to be moving randomly at a high velocity, depending on the temperature used in the calculation. Each atom is influenced by forces such as bonding forces,

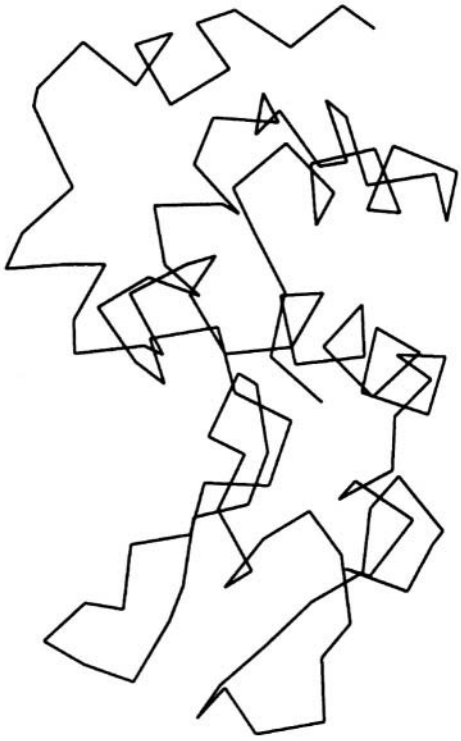


Fig. 9.7. The C^{α} atoms of lysozyme

Fig. 9.8. A cartoon diagram of a protein

angle distortion forces, etc. Since the force on each atom depends on the positions of all the other atoms, and since the positions are continuously changing, the technique is carried out in iterative cycles. First step is to assign the velocities to all the atoms and calculate the initial force felt by each atom in the initial configuration of the molecule. The next step is to calculate the new positions of the atoms after a very short time interval, usually of the order of 0.01 picoseconds (**1 picosecond = 10^{-12}** second). The new positions of the atoms are then used to calculate the new force field. Thus the motion of the various parts of the molecule can be followed. The results are usually checked every 100 steps or so and give 'snapshots' of the molecule at intervals of about 1 picosecond. The simulation usually cannot be carried on for longer than about 5000 picoseconds for the following two main reasons: (a) Enormous amounts of computer time are necessary even on the fastest computers. (b)

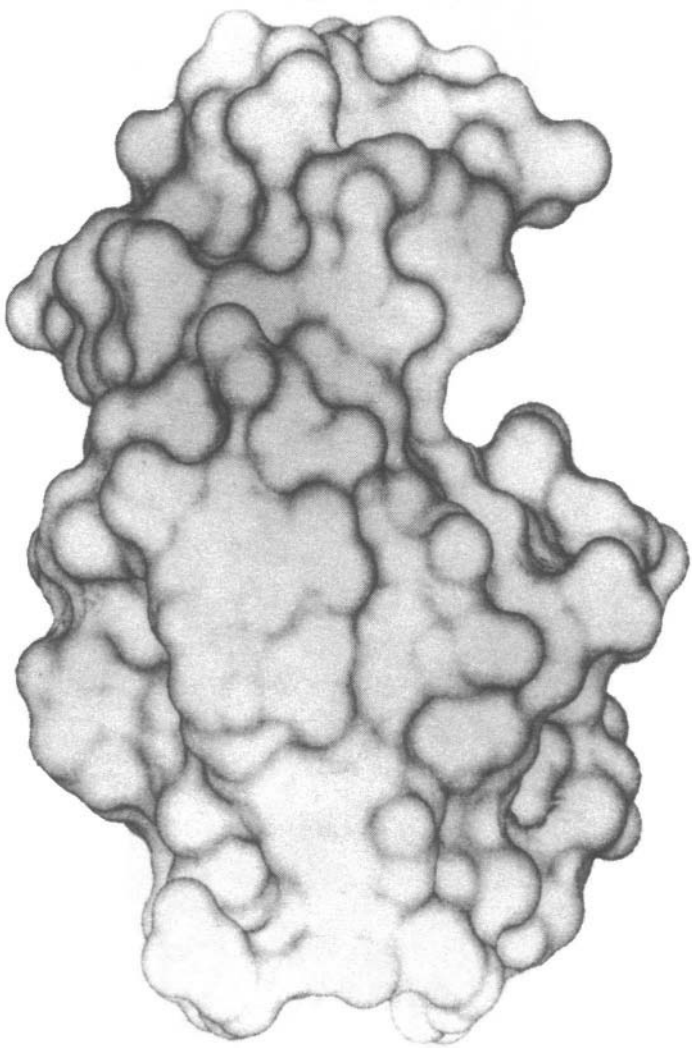


Fig. 9.9. Surface representation of lysozyme.

The force fields are not realistic enough and the molecule tends to assume unrealistic configurations if the simulation is carried on for longer than about 5000 picoseconds. However, large advances have been made in the past five years both in computer technology and in modelling the force-field accurately. This has allowed several interesting simulations of small proteins and DNA fragments to be carried out for as long as 1 millisecond or 10^6 picoseconds. The great advantage of the molecular dynamics technique is that it allows a much more natural picture of the molecule. Molecules are always in motion, but X-ray crystalography and many other techniques give only a static picture.

Apart from building models of molecules, molecular modelling, optimisation and molecular dynamics are used to model interactions between molecules. Successful applications include studying the docking of proteins to substrates, virus assembly, etc. Modelling studies on drug-DNA interactions and drug-protein interactions are used extensively in the rational design of new drugs. Many of the drugs now in the market are the result of such studies.

10

Macromolecular Structure

10.1 Introduction

The central paradigm of biophysics, indeed of physics in general, is

$$\text{Structure} \rightarrow \text{Function}$$

Thus in order to understand and perhaps modify the behaviour of biomolecules, it is necessary to know and understand their structure. One of the most dramatic illustrations of this principle came with the discovery of the double-helical structure of DNA. It was immediately clear how DNA could function as the storage and carrier of genetic information. The vast advances in biotechnology since then stem almost entirely from this single discovery. Protein structure and virus structures also illustrate the truth of this principle. In this chapter we shall consider in detail the structural principles of some the important classes of biological molecules.

10.2 Nucleic Acid Structure

The storage, transmission and expression of genetic information in the cell is carried out by nucleic acids. Deoxyribonucleic acid or DNA is responsible for storing genetic information and passing it on to the progeny cells. DNA is therefore involved in replication and transcription. The other type of nucleic acid present in the cell is ribonucleic acid or RNA. It is involved in many different functions. The genetic information in the DNA is transcribed into messenger RNA or mRNA, which is then translated into the protein in the ribosome. Transfer RNA (tRNA) molecules mediate this process. RNA is also a part of the ribosome and is then called ribosomal RNA or rRNA. Small RNA molecules (~ 100 nucleotides) can fold into structures called ribozymes, which have catalytic activity. During each of these different functions the nucleic acid molecules have different three-dimensional structures.

10.2.1 The chemical structure of nucleic acids

Both DNA and RNA are polymers. The repeating unit, i.e. the monomer in the case of DNA is a deoxyribonucleotide, and a ribonucleotide in the case of RNA. A mononucleotide can be thought as made up of three parts, (a) a phosphate group, (b) a sugar moiety and (c) a nitrogenous base.

The phosphate group is simply PO_4^- as illustrated in Figure 10.1. When it is a part of the polynucleotide, the single extra negative charge is distributed between the two so-called 'pendant' oxygen atoms. It is this negative charge that is responsible for the acidity of nucleic acids. The phosphate group is a rigid unit and has no internal flexibility. The four oxygen atoms occupy the corners of a tetrahedron, with the phosphorous atom at the centre. When the group becomes a part of a polynucleotide, the geometry changes only slightly, with the lengths of the P–O bonds increasing by about 0.2 Å.

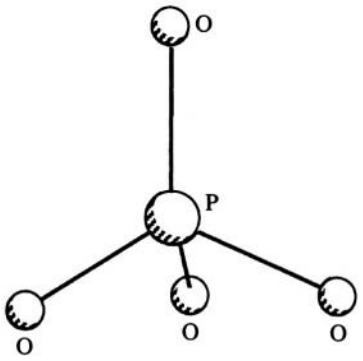


Fig. 10.1 Chemical structure of the phosphate group.

The sugar moiety consists of a five-membered furanose ring with exocyclic alcohol and hydroxyl groups (Figure 10.2). If both the 2′ and 3′ position have a hydroxyl attached, the sugar is ribose (RNA). If there is a hydroxyl group only at the 3′ position it is deoxyribose (DNA). Though it is cyclised, a substantial amount of conformational variation can still take place. Rotation about the C'_4–C'_5 bond produces the three preferred conformations shown in Figure 10.3. In addition, the five-membered ring is not planar, but puckered. Due to the sp^3 hybridised bonds at the carbon atoms of the ring, the energetically favourable conformation is one in which only three or four of the five atoms lie in the same plane. Further, in order to avoid short contacts between the pendant hydrogen atoms, the best way to relieve such energy strains is to push the C'_2 atom or the C'_3 atom out of the plane (Figure 10.4). Thus, in nucleic acids, one comes across C'_2-endo or C'_3-endo sugar puckers, indicating that the displacement of the atom out of the plane of the other four atoms is in the same direction as the C'_5 atom. If the displacement is in the direction opposite to the C'_5 atom, the conformation is termed C'_2-exo or C'_3-exo.

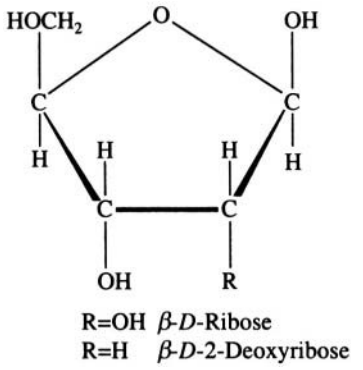


Fig. 10.2 Chemical structure of furanose

The planar bases constitute the chemically variable portion of nucleic acids. Bases can be

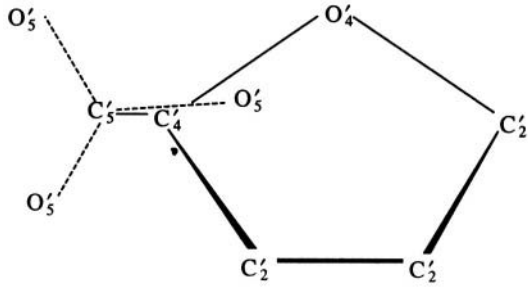


Fig. 10.3 The preferred conformations about the C'_4-C'_5 bond.

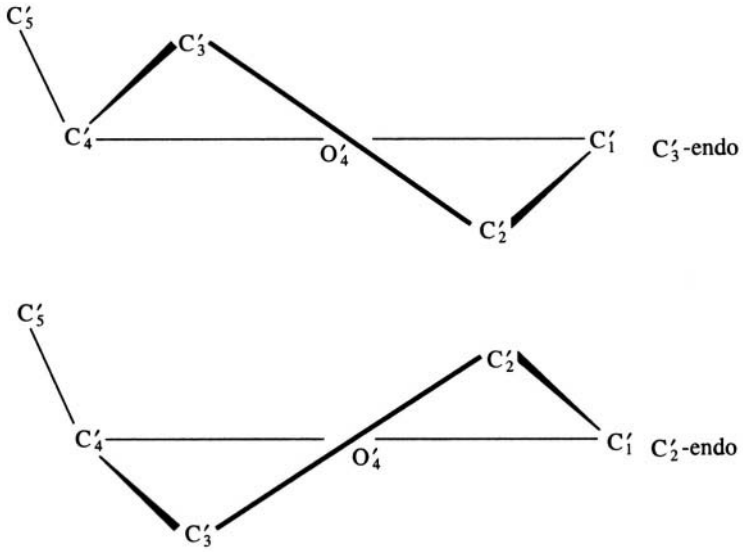


Fig. 10.4 Puckering of the sugar ring.

either pyrimidines or purines (Figure 10.5). The bases guanine, cytosine, adenine and thymine are found in DNA. In RNA the pyrimidine thymine is replaced by uracil. Like the phosphate group, the bases have no internal conformational flexibility.

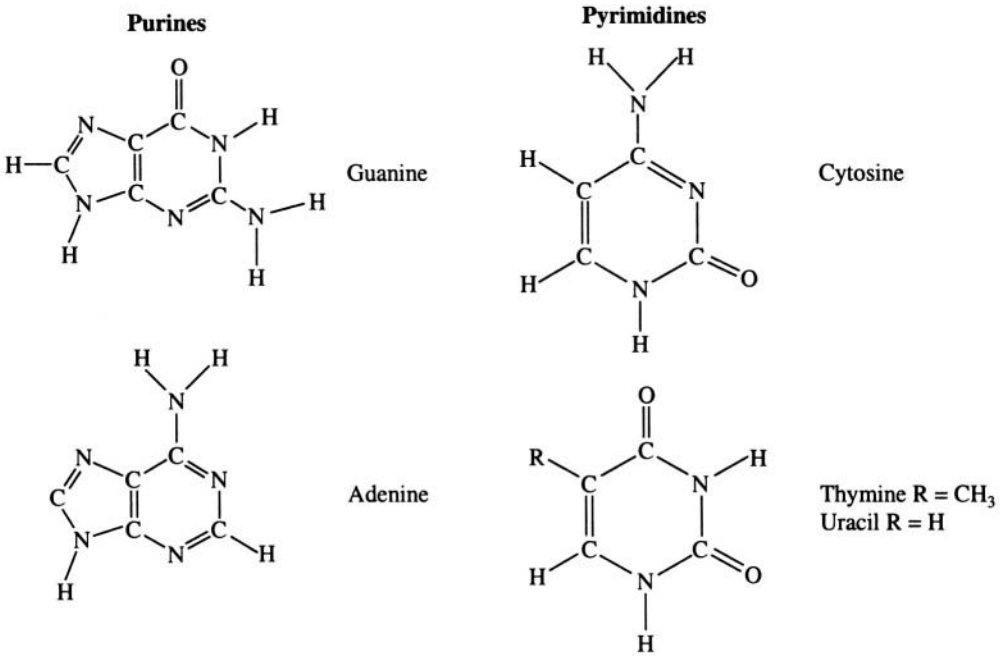


Fig. 10.5 Chemical structure of the nucleic acid bases.

10.2.2 Conformational possibilities of monomers and polymers

A nucleoside is formed by the covalent attachment of one of the bases to a sugar moiety. The bond is called a glycosidic bond and occurs between the C_1' atom of the sugar and either the N_9 atom of a purine or the N_1 atom of a pyrimidine. The plane of the base is almost perpendicular to the approximate plane of the sugar ring. As compared to its components considered separately, a nucleoside has additional conformational possibilities by way of rotation about the glycosidic bond leading to the *anti* or *syn* orientations of the base with respect to the furanose (Figure 10.6).

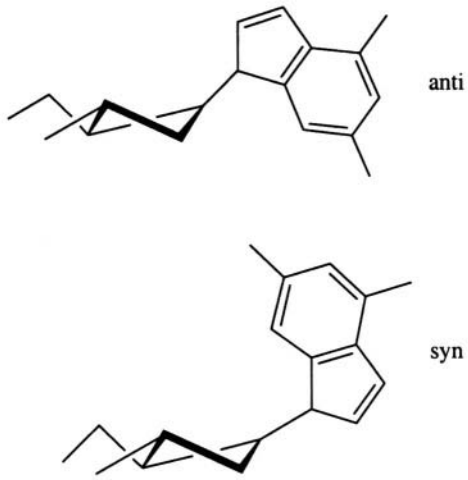


Fig. 10.6 Conformation of the base about the glycosidic bond

The formation of a nucleotide by the addition of a phosphate group to the sugar, at the 5′ end or the 3′ end (Figure 10.7) further increases the conformational possibilities. However, model building, theoretical calculations and experimental evidence from NMR and X-ray crystallography have shown that all conformations are not equally likely.

Mononucleotides link to each other through phosphodiester bonds to form the polynucleotide chain (Figure 10.8). In a dinucleotide there are a total seven bonds about which rotation is possible. In addition, the sugar puckering is also a variable. However, only a few combinations of these angles lead to regular structures when repeated along the chain. Upon formation of the polynucleotides, certain new chemical interactions are introduced which do not exist in the mononucleotide. The most powerful of these are the stacking interactions between the planar bases. These interactions occur between the π-electron clouds and lead to the bases being arranged in parallel planes one over the other like a stack of coins. The inclusion of such

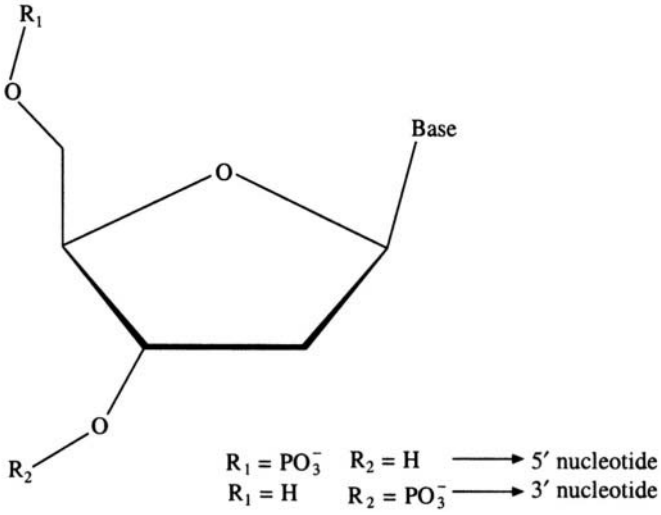


Fig. 10.7 A nucleotide unit

interactions in the calculation of the optimum three-dimensional structure results in a further decrease in the number of possibilities. Some of these structures are discussed below.

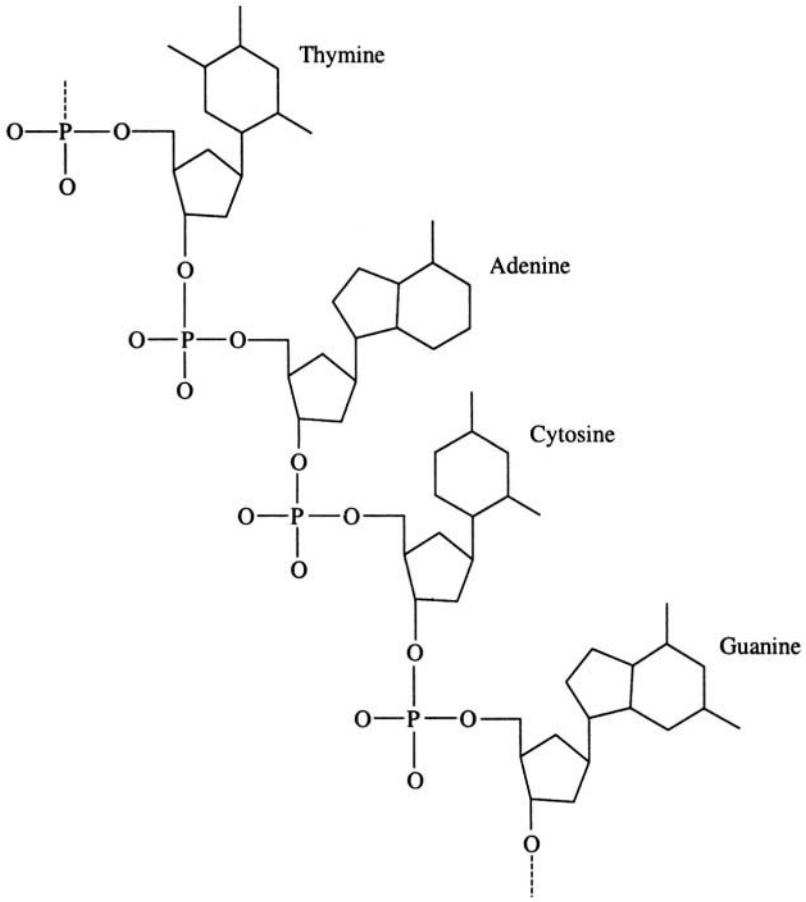


Fig. 10.8. The chemical structure of polynucleotide chain

10.2.3 The double helical structure of DNA

DNA in the cell is one of the largest molecules found in nature. It is a polymer whose length can range from a few thousand nucleotides in bacteria to more than a billion (10^9) nucleotides in cells of higher mammals. The discovery of the double helical structure of DNA in 1953 by J.D. Watson and F.H.C. Crick was one that marked an epoch and is the basis for all of modern biotechnology.

It had been noticed that in the DNA molecules, the number of adenine nucleotides was exactly the same as the number of thymine nucleotides and the number of guanine nucleotides equalled the number of cytosine nucleotides. Furthermore, X-ray diffraction experiments on DNA fibres had suggested that each DNA molecule consisted of two strands which wrapped around each other to form a helix and that these two strands were arranged antiparallel to each other (Figure 10.9). (Note that the attachment of the phosphate group at the 5′ position of the sugar ring is different from its attachment at the 3′ position. This gives rise to a directionality in the polynucleotide chain.) The concept of base pairing explained the base ratio by suggesting that adenine formed specific hydrogen

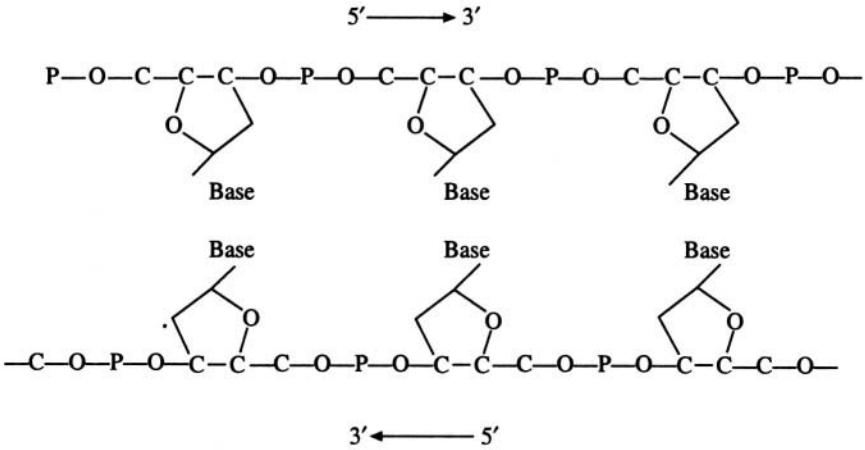


Fig. 10.9 Antiparallel arrangement of the two nucleic acid strands

bonds with thymine and guanine with cytosine (Figure 10.10). Other hydrogen bonding schemes between the bases were much less favourable. Additionally, both the A:T and the G:C base pairs were of the same size. These facts led to the proposal of the structure shown in Figure 10.11. One may visualise the DNA double helix as a flexible rope ladder wrapped around a central pole. The sugar phosphate chains of the two polynucleotide backbones form the sides of the ladder. The base pairs placed perpendicular to the central helix axis form the rungs. The geometrical details of the structure are given in Table 10.1.

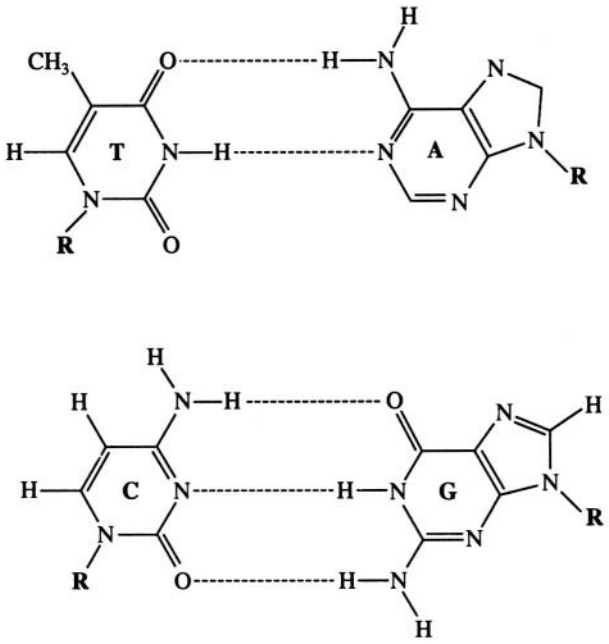


Fig. 10.10 Base pairing in nucleic acids (DNA). In RNA, T is replaced by U

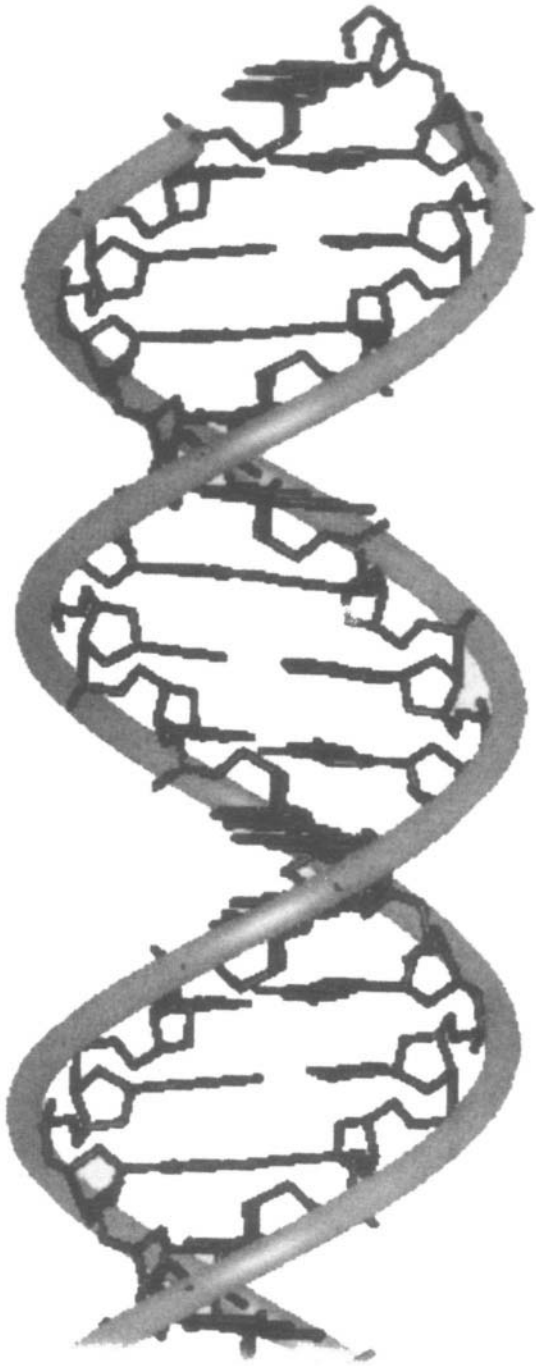


Fig. 10.11 The double-helical, base-paired structure of DNA

Table 10.1 Structural parameters of three polymorphic forms of the DNA double helix

	A	B	Z
Axial rise per residue (Å)	2.56	3.38	3.7
Twist angle per residue (°)	32.7	36.0	–30.0*
Pitch (Å)	28.2	33.8	45.0
Number of base pairs per turn	11	10	12
Handedness of the helix	right	right	left
Number of nucleotides in the repeat unit	one	one	two
Tilt of base pairs from helix axis (°)	20.0	–6.0	–7.0
Dislocation of base pairs into groove (Å)	4.4	–0.14	–5.5#
	minor groove	major groove	major groove
Major groove width (A)	2.7	11.7	8.8
Major groove depth (Å)	13.5	8.5	3.7@
Minor groove width (Å)	11.0	5.7	2.0
Minor groove depth (Å)	2.8	7.5	13.8

*Average of –60° for a dinucleotide repeat. The twist is actually –10° and –50° for the first and second steps respectively of the dinucleotide.
This is an approximate value.
@ This groove is filled up with exocyclic base atoms.

The complementary base-pairing scheme immediately suggests how genetic information is stored and passed on from generation to generation. If the information is 'written' as the sequence of bases, then it may be seen that, when the two strands of the DNA molecule separate, if a complementary strand is built on each of the two strands, it results in two identical copies of the original double helical molecule. This can be repeated again and again and each time the cell divides and in each daughter cell the DNA would be identical to that in the parent cell (Figure 10.12).

10.2.4 Polymorphism of DNA

The structure of DNA proposed first by Watson and Crick explained the X-ray diffraction patterns produced by 'wet' DNA fibres or the 'B' type patterns. This model has thus come to be called B-DNA. Fibres at much lower humidity produced another pattern, with sharper spots and therefore indicating a higher degree of order. They were called the A type patterns and Watson and Crick proposed a somewhat altered model of the double helix to explain them (Figure 10.13). The differences in the helical parameters are quite substantial, as may be seen from Table 10.1. The crucial features that remain unchanged in the two models of DNA are their double helical nature, the antiparallel orientation of the strands and the base pairing scheme. The A type structure also may be imagined as a rope ladder wrapped around a central pole, but the imaginary pole is far thicker in this case. A way of recognising this is to observe the amount of displacement of the base pairs from the helix axis (Figure 10.14). In B DNA, the axis passes through the base pairs, while in A DNA, the axis is displaced by about 2 Å into the major groove. Also the B type model is slimmer than the A type model.

Both A and B type models are right-handed helices in which the helix proceeds as in a right hand screw. A dramatic example of the polymorphism of DNA was discovered in the left-handed Z type DNA (Figure 10.15). It was first observed in sequences of the type 5′ … CGCGCGCG … 3', i.e. alternating pyrimidine-purine. Circular dichroism and UV absorption studies indicated a change in the helical sense from right handed to left handed in sequences of this type. Later a single crystal

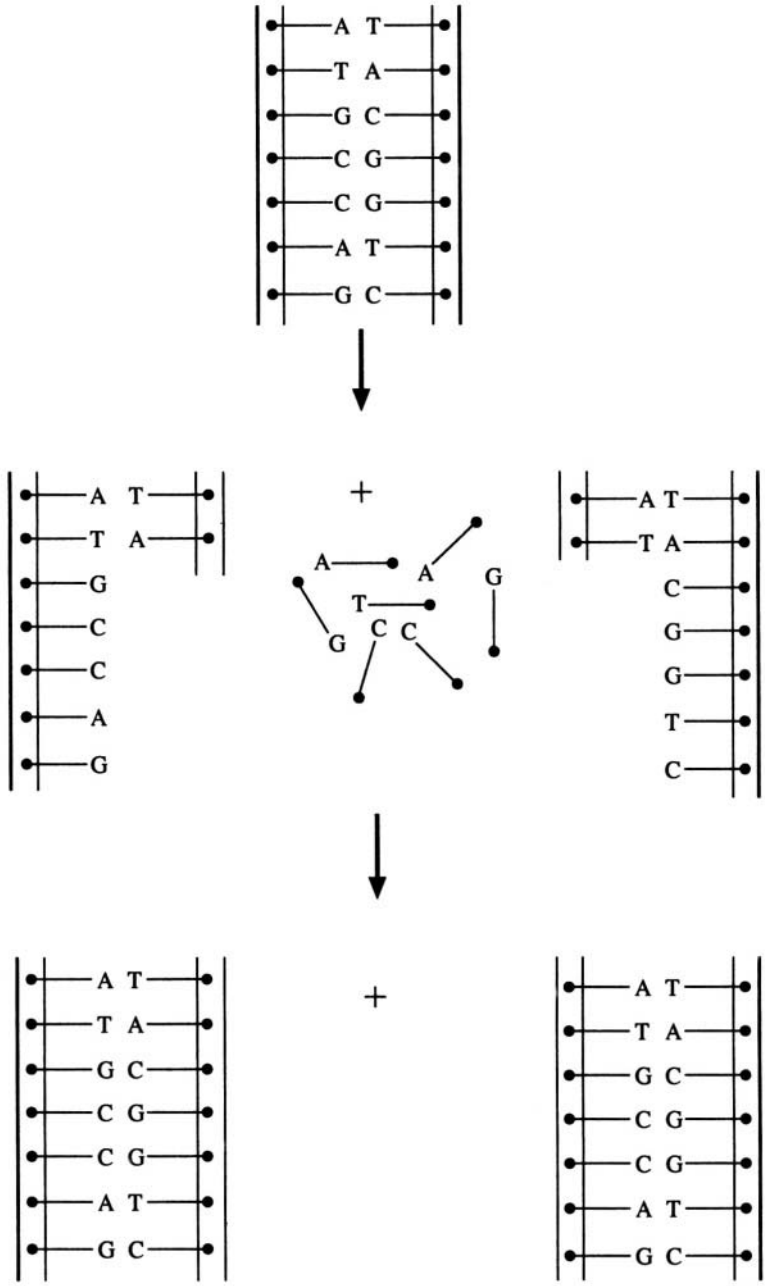


Fig. 10.12 A schematic representation of how the double helix explains DNA replication

structural study of the hexanucleotide 5′-CGCGCG-3′ showed that, not only was the helical sense left-handed but also the phosphate groups followed a zigzag pattern, hence the name Z DNA. The helical parameters of Z type DNA are also given in Table 10.1. It may be noted that left-handed Z DNA has been conclusively and positively identified only *in vitro*. Though antibodies raised against

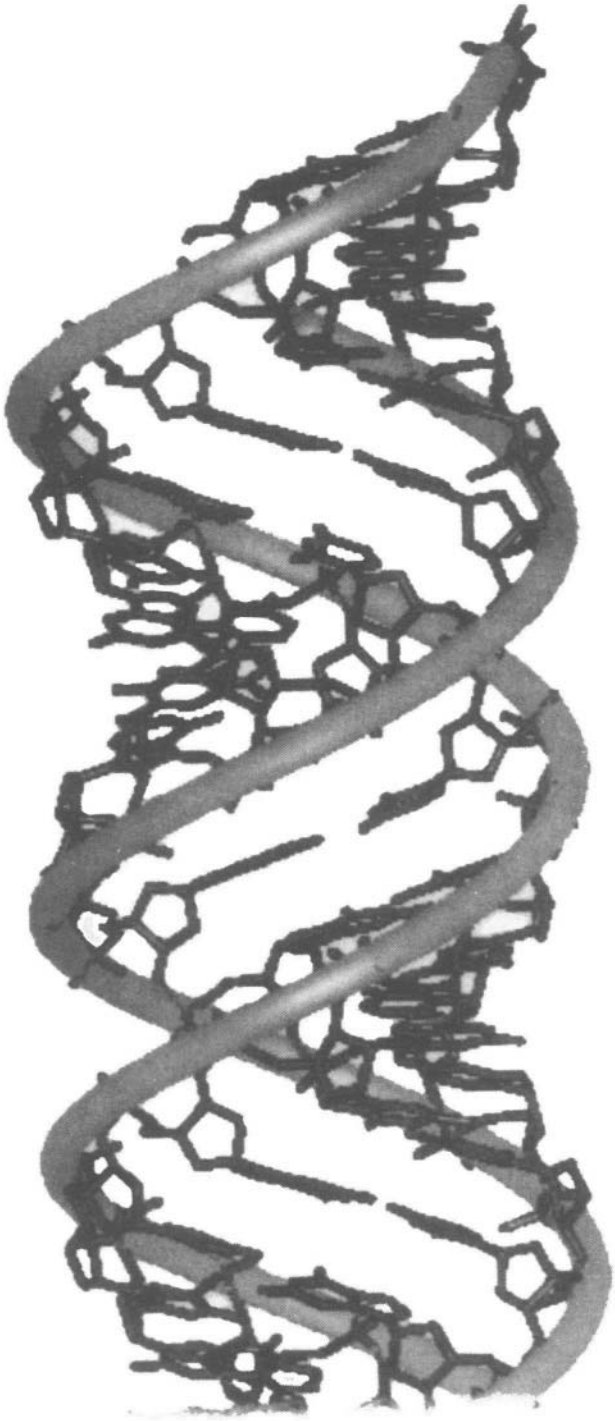


Fig. 10.13 The A type double-helical structure

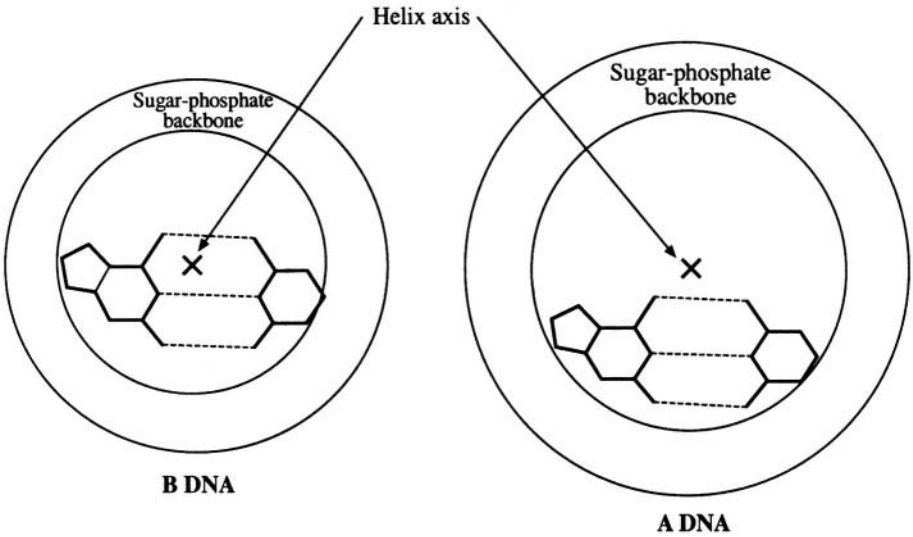


Fig. 10.14 Displacement of the base pairs into the minor groove in A type and B type DNA

such sequences have been shown to cross-react with rat chromosomes, definitive evidence for a biological role for Z DNA is still lacking. Since 1978, single crystal studies of small DNA fragments have yielded high-resolution data on DNA structure. One of the most important findings is that, the structure of DNA deviates from the 'standard' models described above in a sequence dependent manner. The regular models of A, B, and Z DNA, which are described above and in Table 10.1, have been constructed in such a manner that, except for the change in the base, the structure of the molecule, especially the sugar-phosphate backbone, is exactly the same in one part of the molecule as in another. Given the co-ordinates of one monomer and helical transform it is possible to generate the co-ordinates for any length of the double helix.

The crystal structure results go against this and show a microheterogeneity (Figure 10.16). The micro level structure in each portion of the helix appears to depend on the sequence in that part, though the correlation between the sequence and structure is unclear. Sequence specific structure has implications for biology in that it could be another way of transferring information from the DNA molecule (Figure 10.17).

10.2.5 DNA supercoiling and unusual DNA structures

As we have seen, owing to the nature of its stereochemistry, the energetically most favourable state of the DNA molecule is two strands coiled around each other. If, for some reason, the number of times one strand goes around the other is changed from the optimum value, supercoils (Figure 10.18) relieve the extra strain. Exactly the same effect may be seen, for example, in household electrical wires made of two or three twisted strands. In supercoiled DNA, two double helices further wind around each other, like two strands of a rope which are themselves made of coiled fibres. Another way of relieving the strain is by increasing the so-called writhe. In this the DNA molecule wraps round and round an imaginary or real object. In these terms, for example, A type DNA has a higher degree of writhe than B type DNA. The structure of chromatin consists of DNA coils wrapped around histones (Figure 10.19). Each such histone-DNA unit is a nucleosome and in the chromatin state, the nucleosomes are arranged like beads on a string. In its more condensed state, the nucleosome 'necklace' is further wound around to form a solenoid like structure.

Fig. 10.15 Left-handed Z type DNA

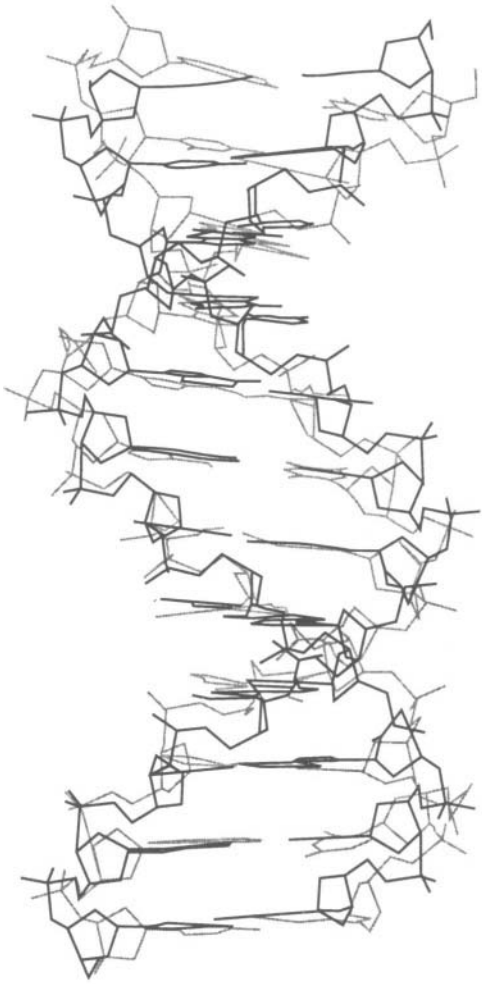


Fig. 10.16 The microheterogeneity observed in the three-dimensional structure of DNA. Dark lines represent the regular helical model derived from fibre diffraction. Light lines correspond to the single crystal structure of the same sequence

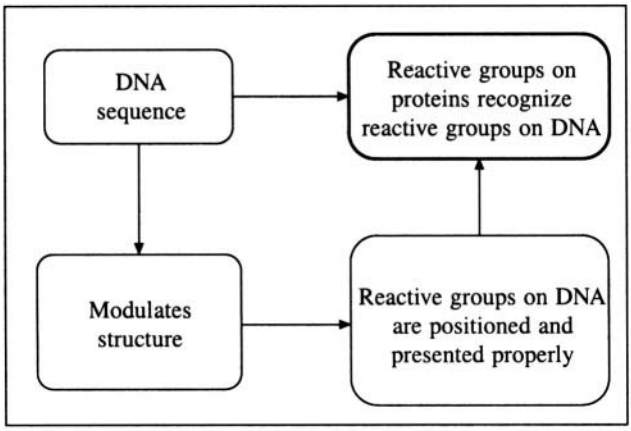

DNA sequence
Reactive groups on proteins recognize reactive groups on DNA
Modulates structure
Reactive groups on DNA are positioned and presented properly

Fig. 10.17 Information transfer from DNA to proteins

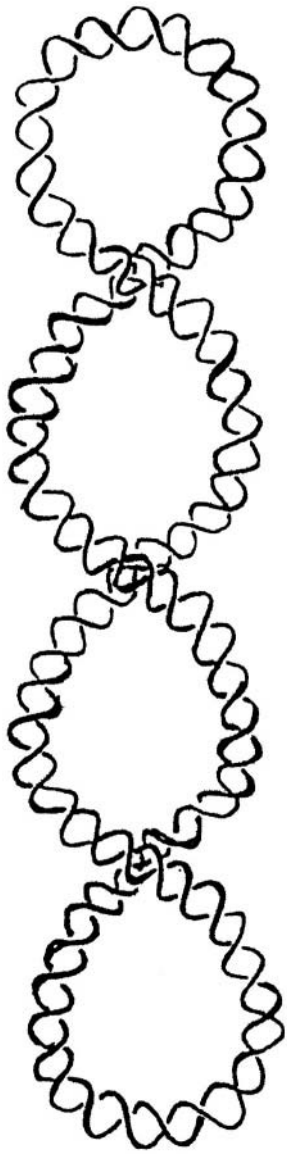


Fig. 10.18 Schematic diagram of supercoiling in DNA

Apart from the double helix, certain DNA sequences have been shown to assume other types of structures. One of these is when stretches of guanine bases come together to form a G-quartet structure (Figure 10.20). Here the guanine bases form hydrogen bonded tetrads that are stacked one over the other. The 'i'-motif is another unusual DNA structure (Figure 10.21) formed by cytosines and which has two distinctive features: (i) The cytosine-cytosine base pair. (ii) The intercalated structure of the two duplexes which make up the tetraplex.

Other unusual DNA structures, observed using various biophysical techniques, include triplexes (Figure 10.22), loops, cruciforms, slipped DNA, etc.

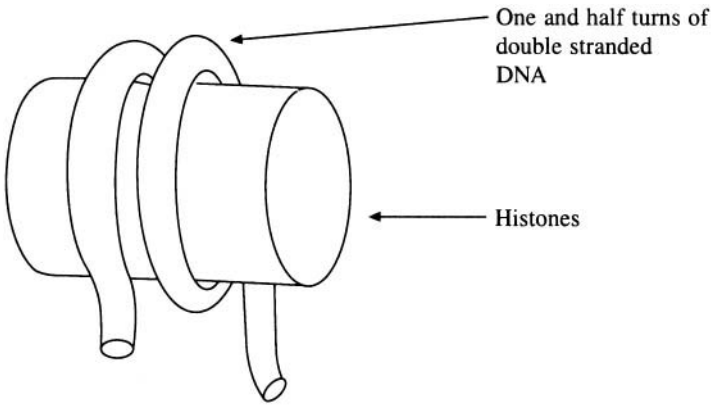


Fig. 10.19 The structure of a nucleosome core particle

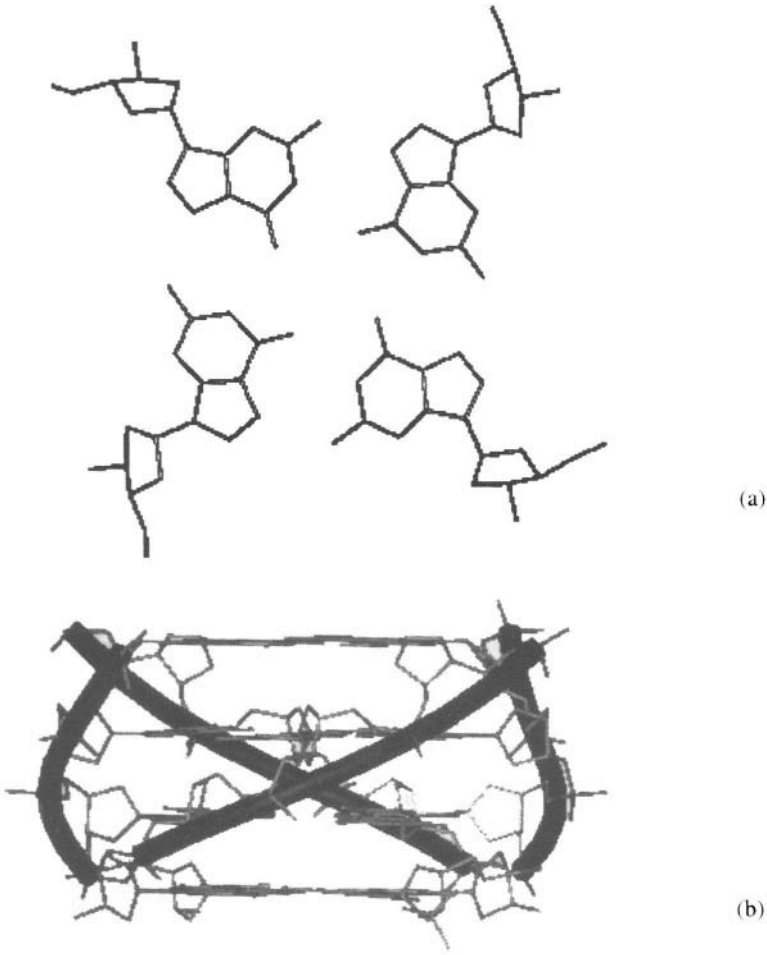


Fig. 10.20 Structure of G-plates. (a) View down the helical axis showing the arrangement of four guanine residues in a plane. (b) View perpendicular to the helical axis

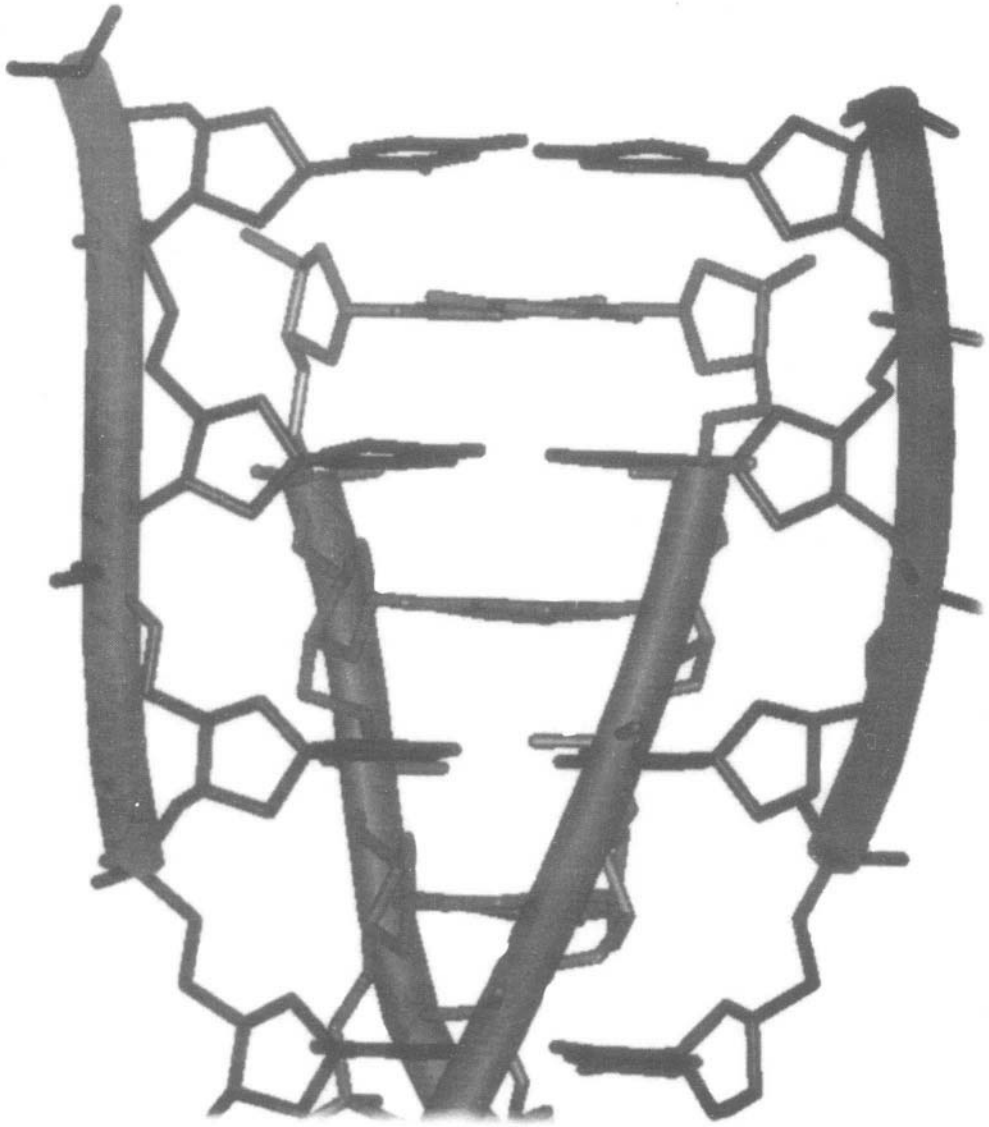


Fig. 10.21 The 'i'-motif structure

10.2.6 The structure of transfer RNA

Most of the ribonucleic acid in the cell exists as rather short single strands, compared to the very long double stranded form assumed by DNA. For example, transfer RNA (tRNA) molecules are only about 70-80 nucleotides long. Some of these nucleotides consist of unusual bases such as psuedouridine and dihydrouracil. Messenger RNA (mRNA) molecules have varying lengths, depending on the code they carry. Ribosomal RNA (rRNA) molecules, which form a part of the ribosome, are also long nucleotides.

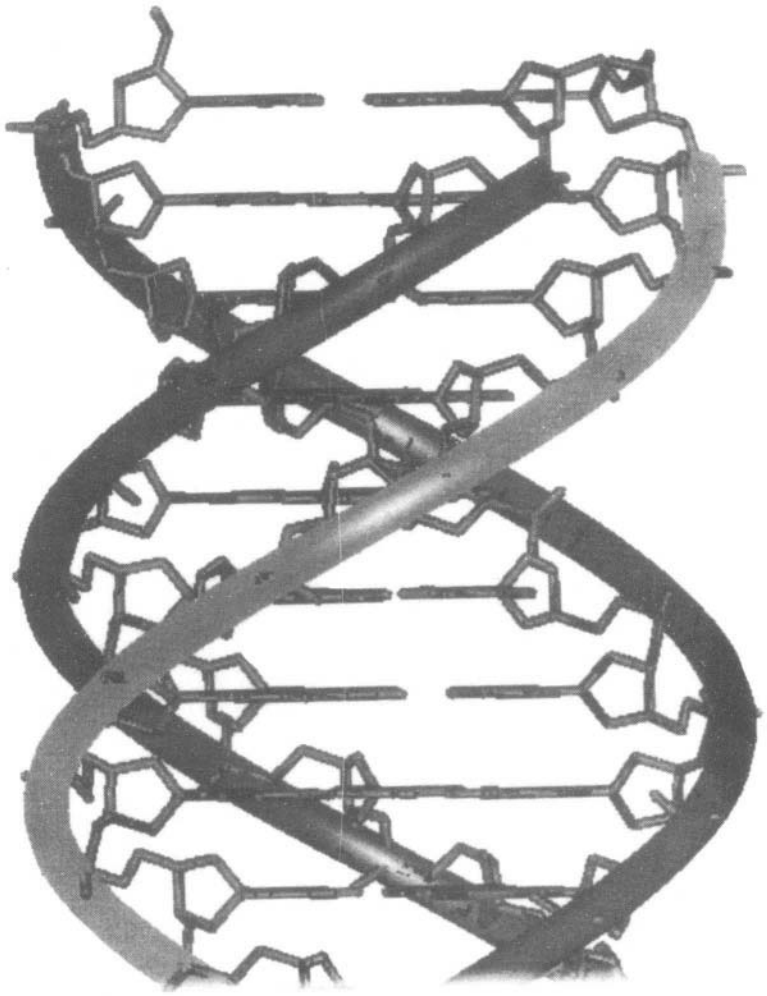


Fig. 10.22 The structure of triple-stranded DNA.

Only the structure of tRNA is known in some detail. Being a single strand, the structure is not a double helix. However portions of the sequence can form Watson-Crick base pairs with other portions of the molecule and these regions are double helical. An analysis of the sequences of tRNAs suggests that they can be arranged to form four base-paired regions so that the general shape is that of a clover leaf (as seen in playing cards) (Figure 10.23). Earlier theoretical and modelling studies, supported by diffraction and other physico-chemical studies, had shown that RNA double helices must necessarily be of the A type. This is because steric hindrances due to the extra oxygen atom at the 2′ position of the sugar which would prevent the other structural types of helices. In tRNA too, the base paired stems form short A type helices. The three dimensional arrangement of these short helices and the extra loops is such that the structure looks like the letter 'L' (Figure 10.24). The anticodon region, i.e. the portion that carries the code is at one end and the amino acid is attached at the other end.

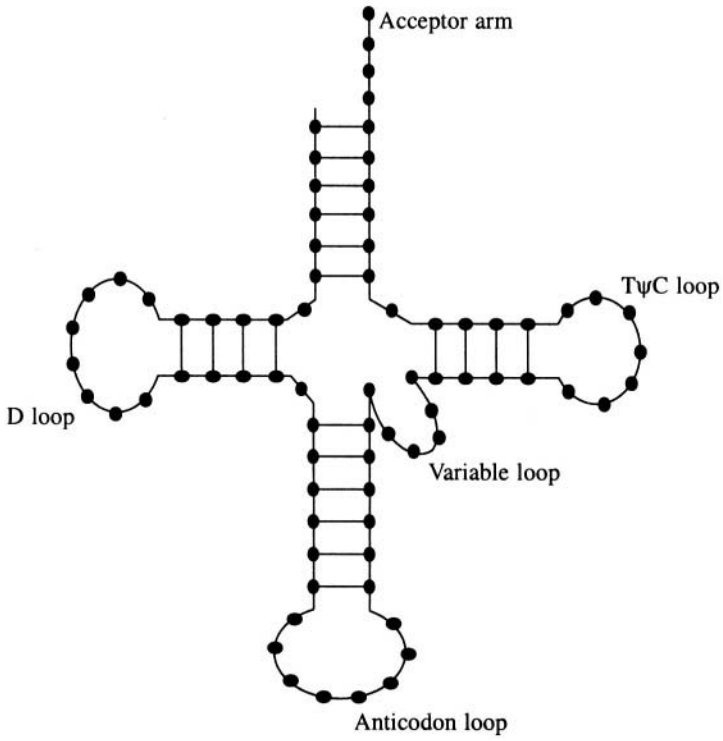


Fig. 10.23 Clover leaf secondary structure of tRNA.

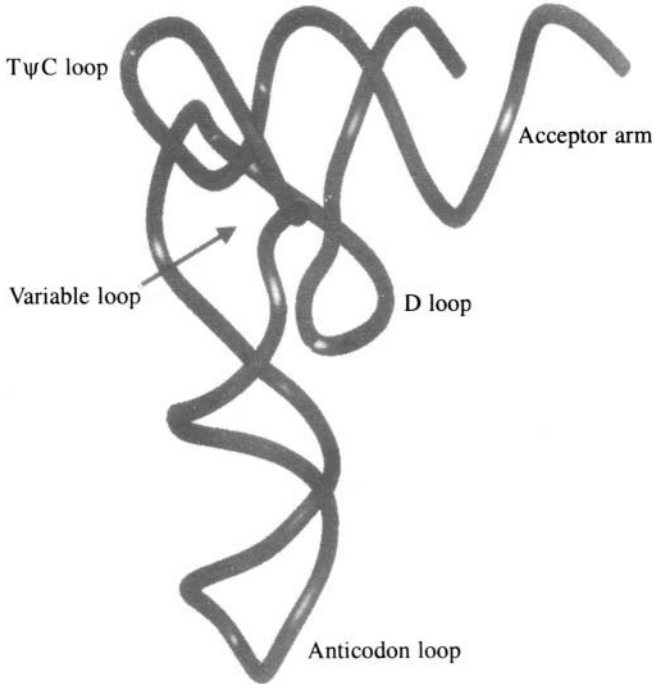


Fig. 10.24 Three-dimensional structure of tRNA. Only the trace of the sugar-phosphate backbone is shown.

10.3 Protein Structure

Proteins are an important class of biological polymers. They are the molecules that actually do almost all the work in the cell and are largely responsible for expressing the information present in the DNA molecule. They range in size from a few kilodaltons to hundreds of kilodaltons. All proteins are built up from 20 amino acids, which constitute the monomers. The sequence in which these amino acids are arranged differs from protein to protein and is referred to as the primary structure of the protein. Some sequences of amino acids are known to arrange themselves into regular three-dimensional structures that are referred to as the secondary structure of the proteins. At the next level of structural organisation, the secondary structural units arrange themselves into globular shapes. This is called the tertiary structure and is stabilised mainly by interactions between the amino acid side chains. Thus the side chain conformations of the amino acids also form part of the tertiary structure. Also discernible at this level are domains corresponding to compactly arranged groups of secondary structural features. In the next higher level of structure, globular folded polypeptide chains come together in specific arrangements called the quartenary structure. Thermodynamic experiments by Anfinsen have shown that the functional three-dimensional structure of the protein is encoded in its primary amino acid sequence. Thus a polypeptide chain in solution will automatically fold into its three dimensional structure given only that the conditions of temperature, ionic strength, pH, etc., are within a fairly wide range. Ultimately the structure and therefore the function of proteins is directly derived from the chemical nature of the amino acids.

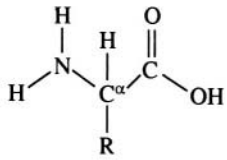
10.3.1 Amino acids and the primary structure of proteins

Alpha amino acids can be considered as being chemically put together around the central carbon atom called C^{α}. Carbon atoms have a valency of four and in amino acids this valency is satisfied by an amino group, a carbonyl group, a hydrogen atom and a side chain. The only exception relevant to proteins is the amino acid proline. Figure 10.25 gives details of the side chain of the twenty amino acids that are found in naturally occurring proteins. It is not known why only these twenty amino acids were chosen. Presumably the answer lies in the way in which life originated on earth.

The 20 amino acids have two major features in common (a) they are all α- amino acids and (b) they are all L-amino acids. They can be separated into groups with other common features. Glycine, alanine, valine, isoleucine, leucine and phenylalanine have hydrophobic, non-polar side chains. Phenylalanine, tryptophan and tyrosine have aromatic planar rings. Aspartic acid and glutamic acid carry a negative charge at neutral pH, while lysine and arginine each carry a positive charge. Histidine has a highly reactive 5 membered planar ring that also carries a positive charge at neutral pH. Cysteine, serine, threonine, asparagine, glutamine and tyrosine have neutral, polar (and therefore hydrophilic) side chains. The chemical nature and size of the side chain determines its activity and also its position in the protein. Histidine frequently occurs at the active site of enzymes. Hydrophobic side chains usually occur in the interior and hydrophilic groups on the surface of the proteins that are active in aqueous media. The situation is the reverse in the case of membrane proteins. The sequence in which the amino acids occur along the polypeptide chain therefore closely controls how the chain will fold into the three-dimensional structure. In other words the primary structure of the protein determines its secondary, tertiary and quaternary structures and ultimately the function of the protein. The polypeptide chain is built up when the amino acids join to each other by means of a peptide bond.

10.3.2 The peptide bond and secondary structure of proteins

The peptide bond is formed as shown in Figure 10.26. The C-N bond has a partial double bond character, which means that the molecule cannot rotate about it. The six atoms that form the peptide



The general formula of an amino acid
R = sidechain

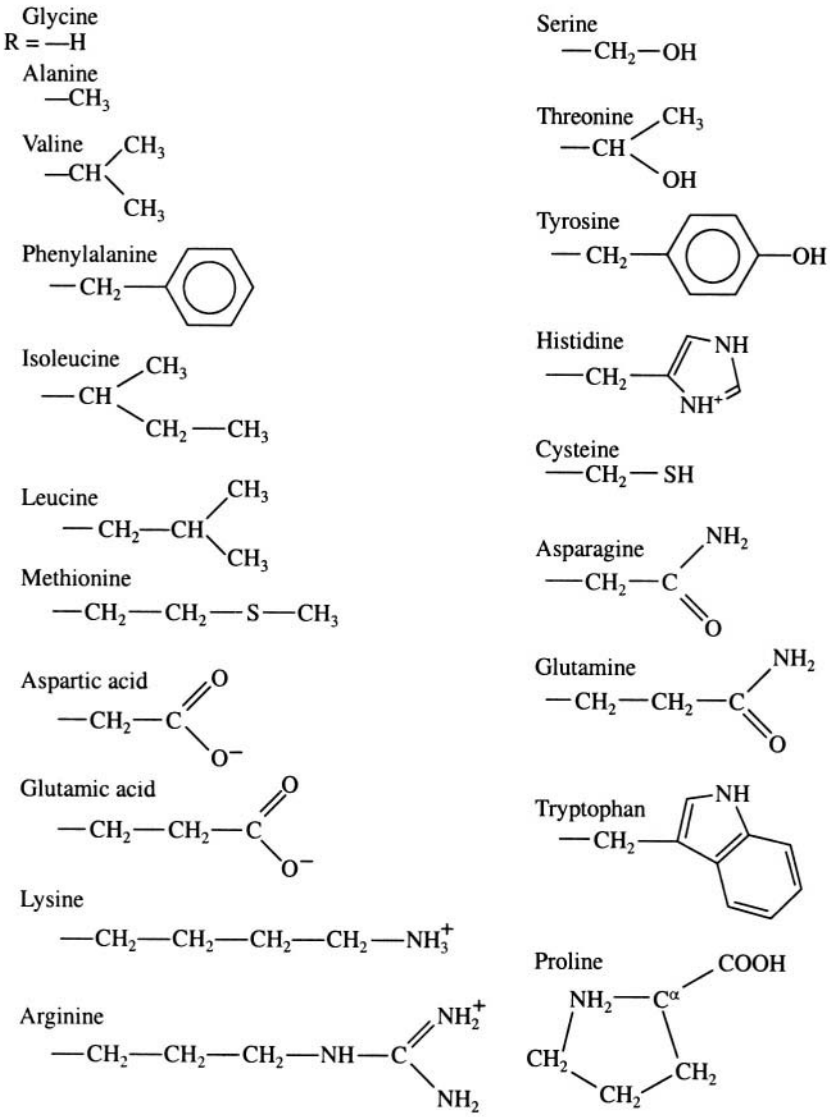


Fig. 10.25 Chemical structures of the alpha-amino acids

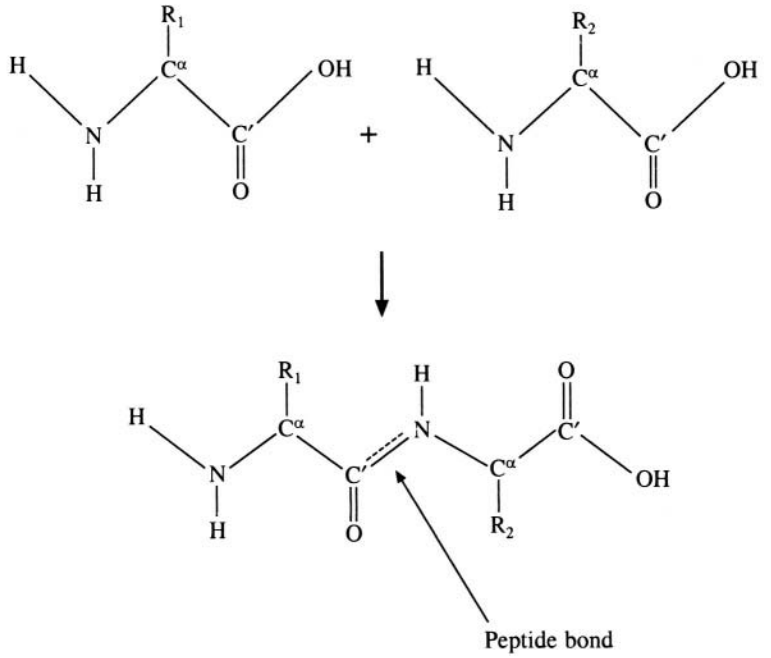


Fig. 10.26 The peptide bond. Dotted line indicates a partial double-bond character.

group are constrained to lie on a plane thereby forming the planar peptide group (Figure 10.27). Ramachandran and his colleagues at the University of Madras realised that this property simplified the geometrical analysis of the polypeptide chain.

If the rotational possibilities of the longer side chains are ignored and the backbone of the protein chain alone is considered, then each residue is described by two torsion angles ϕ and ψ (Figure 10.28). The freedom of rotation about even these bonds is not absolute, but is restricted by possible steric hindrances. For example, if the $C'_{i-1} - N_i$ bond is *cis* to the $C_i^{\alpha} - C'_i$ bond when the $C_i - N_{i+1}$ bond is *cis* to the $N_i - C_i^{\alpha}$ bond, i.e. when both ϕ and ψ are 0°, then a severe clash between H_{i+1} and O_{i-1} will occur, making this particular pair of values disallowed. A systematic search of all pairs of values of ϕ and ψ reveals that, only about 18% of the values are allowed. This is shown as the Ramachandran map (Figure 10.29). This map indicates the situation when the atoms are modelled as hard spheres allowing no interpenetration. The radii of the spheres were derived from the interatomic distances seen in the crystal structures of small molecules. Apart from these normal values, somewhat smaller values for the contact distances were also obtained. These were called the extreme limits, and when they were used, the allowed area on the Ramachandran Plot expanded to 22%. The map is correct for eighteen of the twenty side

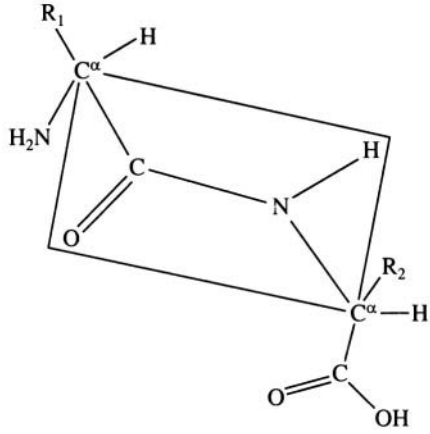


Fig. 10.27 The planar peptide unit

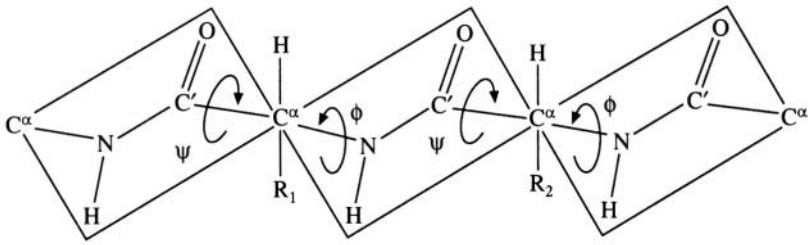


Fig. 10.28 The torsion angles ϕ and ψ used to describe the conformation of a polypeptide chain

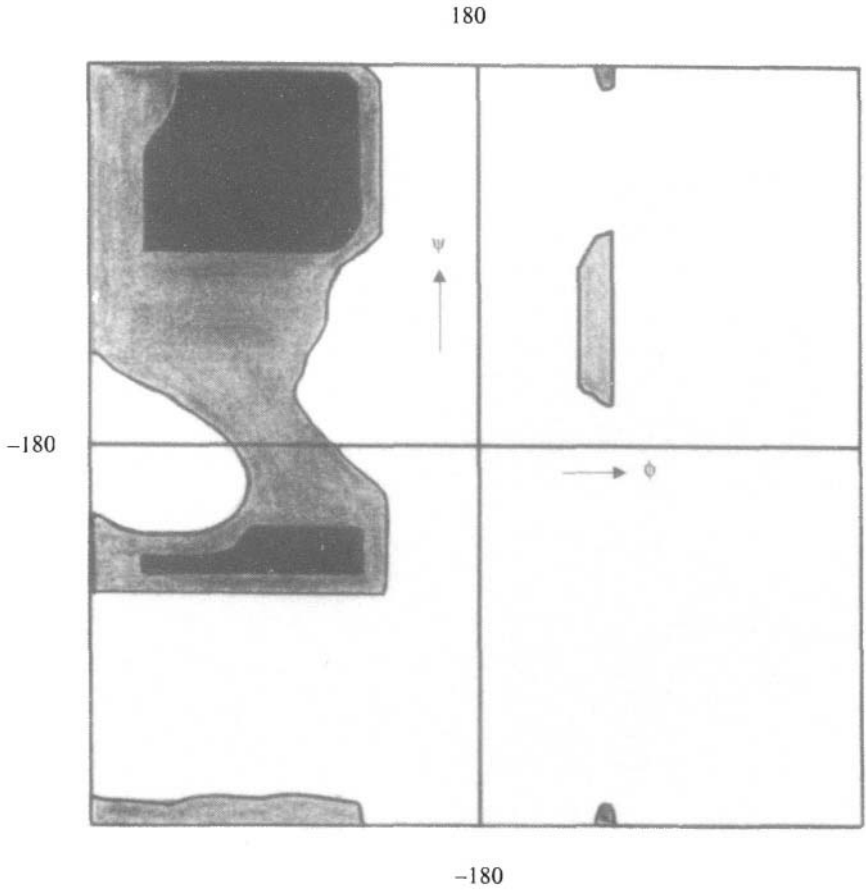


Fig. 10.29 The Ramachandran map. Strictly allowed regions are black while generously allowed are light grey.

chains. The exceptions are glycine and proline. Glycine has a much larger allowed region owing to the absence of the C^{β} atom (Figure 10.30). The side chain of proline binds to the peptide nitrogen atom and this limits the value of ϕ to $-60 \pm 20°$.

The peptide units are not absolutely planar, as assumed in constructing the above maps. The torsion angle ω about the peptide bonds can deviate by 10 to 20° from planarity. This changes the ϕ–

ψ map somewhat. The map also changes when the 'rigid sphere' model of the atoms is replaced by a semiempirical energy function involving interatomic distances. The changes are however quite small and over 98% of the experimentally determined values of ϕ and ψ fall within the allowed regions of the Ramachandran plot.

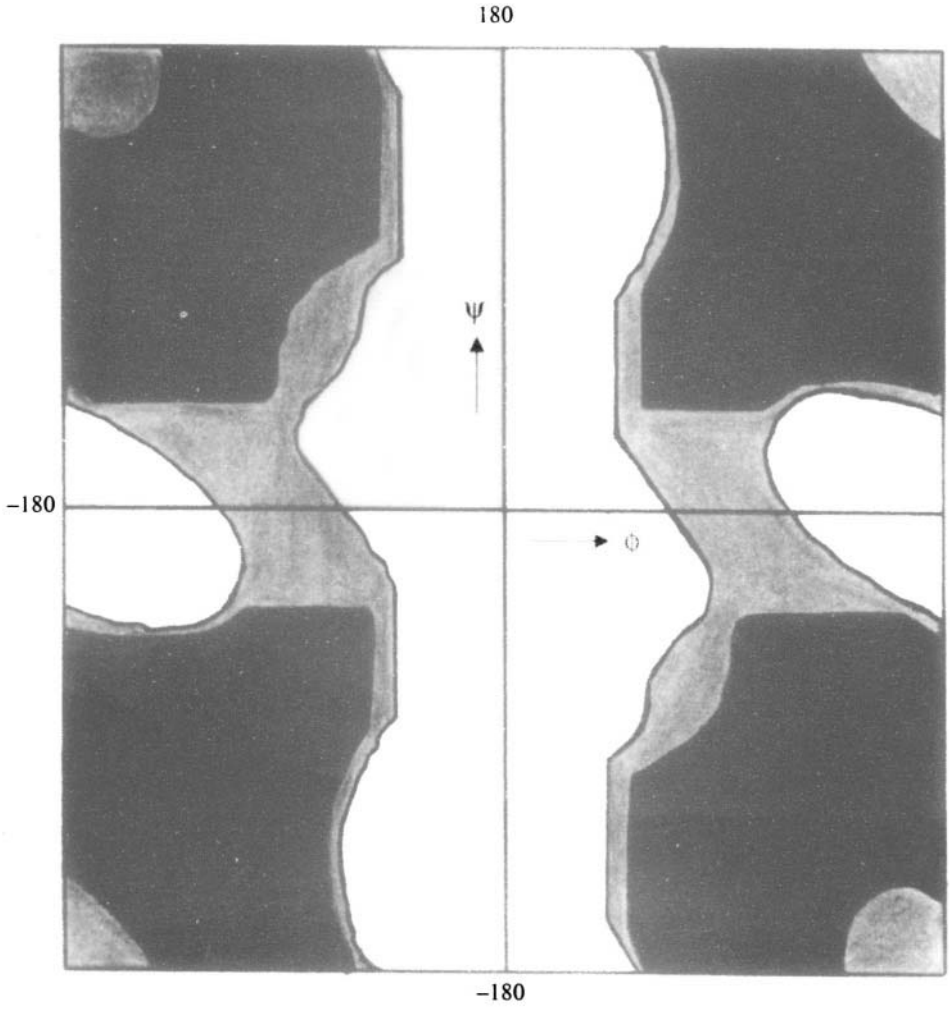


Fig. 10.30 The Ramachandran map for glycine. Strictly allowed regions are black while generously allowed are light grey.

If some of the pairs of values of ϕ and ψ which fall within the allowed region are repeated for many residues along the polypeptide chain, it leads to regular recognisable secondary structures, such as **α-helices** and **β-sheets.**

10.3.2.1 ***α-helix*** This is one of the most commonly occurring secondary structures in proteins, with (ϕ, ψ) values of (–57°, –47°) (Figure 10.31). The helix has 3.6 residues per turn. The occurrence of such a non-integral value for the helical repeat was a surprise when Linus Pauling first proposed it.

But subsequent experimental and theoretical studies have confirmed the structure overwhelmingly. The stability of the **α-helix** is enhanced by the hydrogen bonds between the C=O group of residue '*i*' and the NH group of residue '*i* + 4'. Each turn of the **α-helix** is 5.4 Å long, indicating a rise per residue of 1.5 Å. The width of the helix is about 4 Å. For *L*-amino acids only right handed **α-helices** are allowed and therefore almost all **α-helices** observed in protein structures have a right handed twist. Very short (3-5 residues) left-handed **α-helical** regions have also been observed in proteins, though very rarely. Owing to the fact that all the hydrogen bonds in the **α-helix** point in the same direction, the helix as a whole has a dipole moment such that the amino end or the *N* terminal has a net partial positive charge and the carboxy end or the *C* terminal has a net partial negative charge. This induces negatively charged ions such as PO_4^- to bind to the amino end. However, positively charged ions rarely bind to the carboxyl end. Alpha helices are preferentially made up of certain of the 20 amino acid residues. Among the strongest helix-formers are alanine, glutamine, leucine and methionine. Conversely proline, glycine, tyrosine and serine occur in helices only rarely. An **α-helix** often occurs on the surface of a protein. Then it is usually made up of hydrophobic residues on the side facing the interior of the protein and hydrophilic residues on the other.

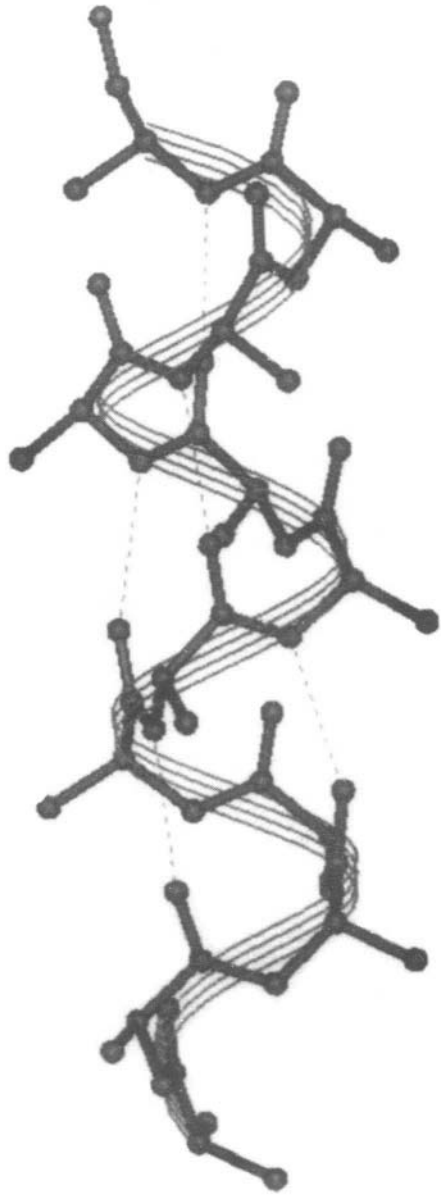


Fig. 10.31 The alpha helix

*10.3.2.2 **β-strands** and **β-sheets*** **β-strands** (Figure 10.32) are highly extended polypeptide chains and the value of (ϕ, ψ) are in the large allowed region in the top left quadrant of the Ramachandran Map. **β-strands** interact with other **β-strands** to form **β-sheets**. This interaction can occur in two ways. Firstly, all the interacting **β-strands** could run in the same direction, i.e. the amino terminii of all strands are on the same side. In this case, a parallel a **β-sheet** is formed (Figure 10.32(a)) in which somewhat distorted H-bonds are formed between the amino groups of one strand and the carboxy groups of the other. In the other arrangement, the strands run in opposite directions, resulting in an antiparallel sheet (Figure 10.32(b)). Here again N–H … O hydrogen bonds stabilise the structure. In both cases, the amino acid side chains project out of the sheet on either side. The sheets are usually not flat but have a twist (Figure 10.33) and always in a right handed sense. Mixed parallel and antiparallel **β-sheets** also occur, but only rarely.

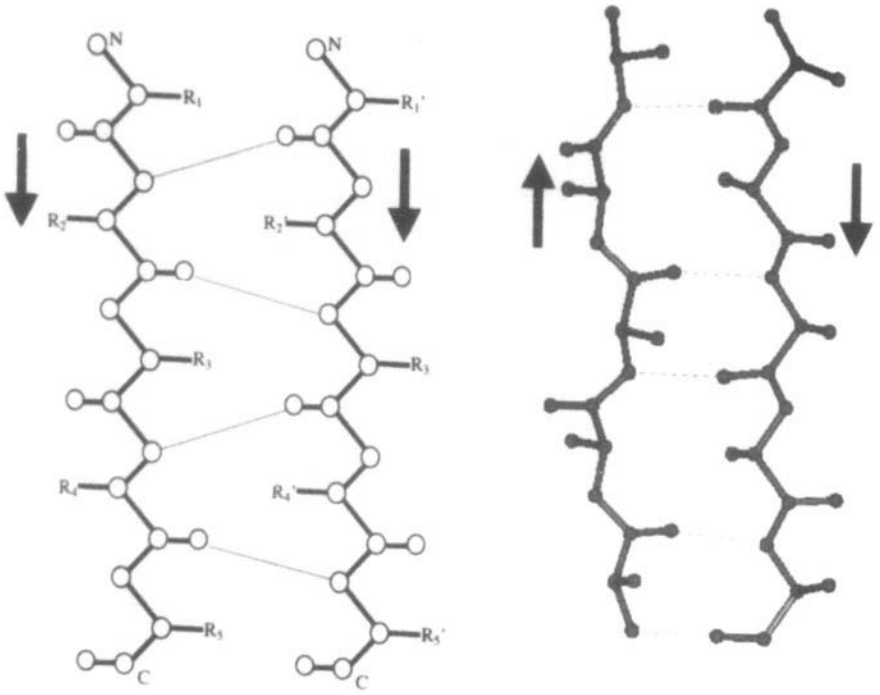


Fig. 10.32 Beta sheets: (a) parallel (schematic) and (b) anti-parallel (3-D representation)

10.3.2.3 ***β-turns*** Whenever the peptide chain has to reverse its direction, as frequently happens in globular proteins, it can sharply and efficiently do so through a structure known as the **β-turn** or a reverse turn. Three consecutive peptide units are involved in such a turn and the most prominent stabilising factor is the hydrogen bonds that are formed between O_i and N_{i+3} (Figure 10.34). The (ϕ, ψ) values of all the residues involved in this secondary structural unit do not fall in the same region of the Ramachandran plot. At the residue i+1, the values are (–60, –30°) and at the residue i+2, the values are (–90,0°), assuming that the turn is made up of the residues i, i+1, i+2 and i+3. This is called the Type I **β-turn.** In the Type II turn, the values are (–60,120°) and (80,0°) at residues i+1 and i+2 respectively. However, in this turn there is a steric clash between side chain R_{i+2} and O_{i+1}, so that R_{i+2} can only be H, i.e. glycine. A type III reverse turn has regular structure with (ϕ, ψ) values of (–60, –30°) at both central residues. A repetition of these angles over a long stretch leads to a type of helix called a 3_{10} helix, which is found in proteins, though rarely.

10.3.2.4 Collagen helix This is a special type of helix discovered by Ramachandran and his colleagues. It occurs in the fibrous protein collagen, which is the most abundant protein in animals. It is a left-handed helix, with (ϕ, ψ) values about (–60,140°). It is not a completely regular helix and the (ϕ, ψ) values are repeated exactly only at every 3 residue, i.e. at the glycine residue (the sequence of collagen is always (gly-x-y) where x and y are frequently proline and hydroxyproline respectively). The protein collagen itself consists of three of these single strand helices wound around each other to form a triple helical fibre (Figure 10.35). The structure was able to explain why such a repetitive sequence was required, since every third residue came close to the superhelix axis. Side chains larger than that of glycine would introduce a large degree of instability.

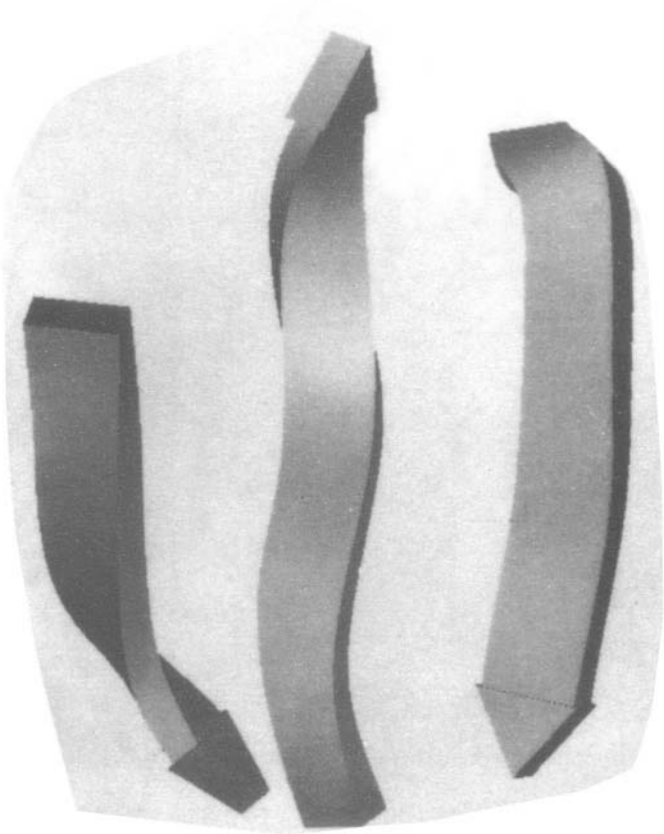


Fig. 10.33 The twist in the beta sheet.

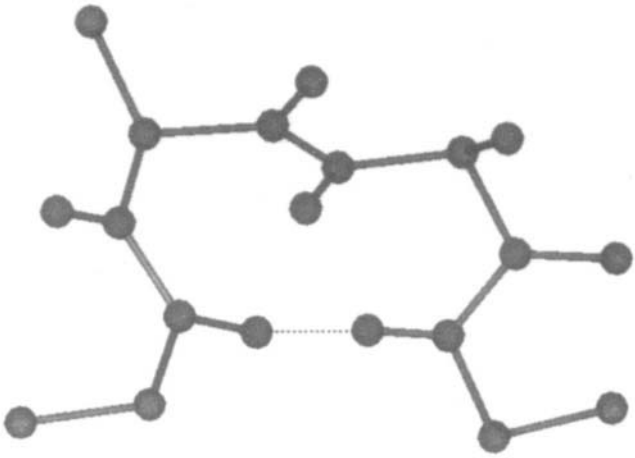


Fig. 10.34 A beta turn

10.3.3 Tertiary structure-supersecondary and domain structure

Various segments of a protein chain fold into various secondary structural units such as **α-helices** or **β-sheets.** These units further fold up into compact supersecondary structural units or domains. This is called the tertiary structure of the proteins and involves interactions between amino acid residues that are not necessarily close together in the primary sequence. One of the most important of such tertiary interactions is the disulphide bond.

The disulphide bond is formed between the sulphur atoms of two cysteine residues (Figure 10.36). Such bonds could occur within a single polypeptide chain or between two different chains. The chief function of the disulphide bridge is to provide mechanical stability to the protein structure. They may also determine chemical properties by stabilising the correct active conformation. In some proteins the disulphide bond may play a catalytic role.

Another major stabilising factor in the tertiary structure of proteins is the so-called entropic or hydrophobic forces. Since the protein is made up of both polar and non-polar side chains, in an aqueous solution the protein molecule behaves like a drop of oil, with polar groups on the surface, in contact with the solution and the non-polar groups in the interior of the molecule. The solvent molecules in the near vicinity of the protein also assume a certain amount of ordering around polar groups. For non-polar groups, however, the solvent molecules make a large negative contribution to the entropic energy, leading to the folded state being favoured over the denatured state. Other factors which stabilise the tertiary structure of proteins include salt bridges or strong Coulombic interactions between oppositely charged amino acids, such as glutamic acid and aspartic acid. Hydrogen bonds in the interior of the protein are another important stabilising factor not only of the secondary structural motifs but also of the tertiary structure.

The secondary structural units often form clearly identifiable domains or super-secondary

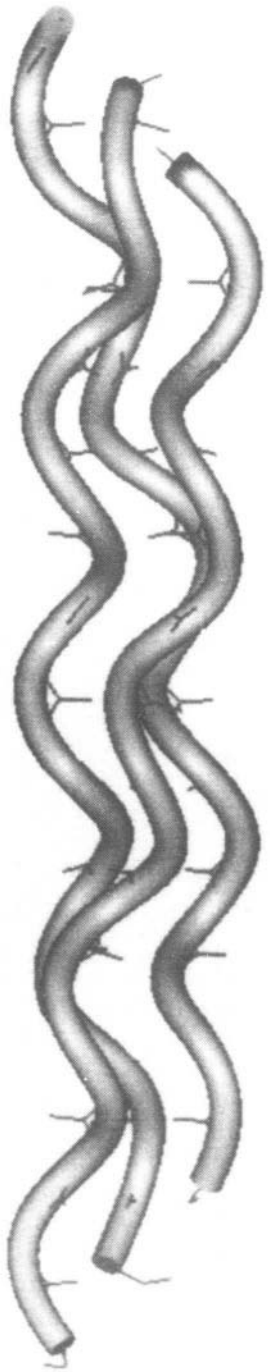


Fig. 10.35 Triple-helical structure of collagen

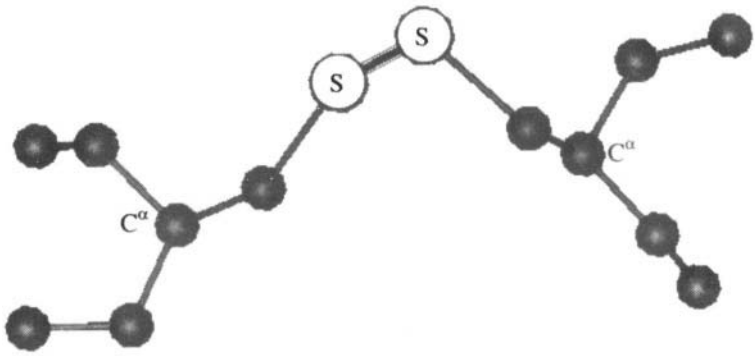


Fig. 10.36 The disulphide bond.

structures. In proteins like cytochrome b_{562}, four helices come together to form the so-called 4 helix bundle (Figure 10.37). This is a widely occurring super-secondary structure. Motifs containing both **β-sheets** and **α-helices** also occur in many proteins. An example is α/β barrel structure. It has been noticed that the core of such barrels is usually packed with hydrophobic side chains. Other super-secondary structural motifs include **β-barrels** such as the one seen in retinol binding protein. Domains in proteins can be loosely defined as an almost self-contained globular portion of the protein, loosely connected to other domains to form the complete molecule. A clear example of a domain structure is the protein troponin C (Figure 10.38). In other cases the domains may not be so clearly demarcated. However, calculations of the distances between pairs of residues can be performed and displayed on a so-called diagonal plot, which would clearly reveal structural domains in proteins.

10.3.4 Quaternary structure

Many proteins in their active or functional forms exist as aggregates of more than one folded polypeptide chain. The macromolecular structure so built up is called the quaternary structure of proteins. Usually the subunits of quaternary structure associate by non-covalent interactions, though disulphide linkages may exist between different subunits. The monomers, which form the quaternary structure, may be identical or may

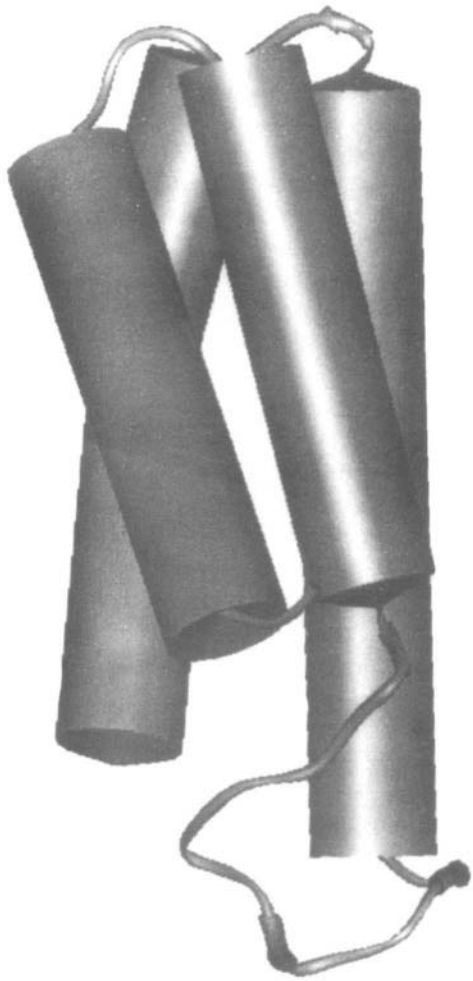


Fig. 10.37 A four-helix bundle.

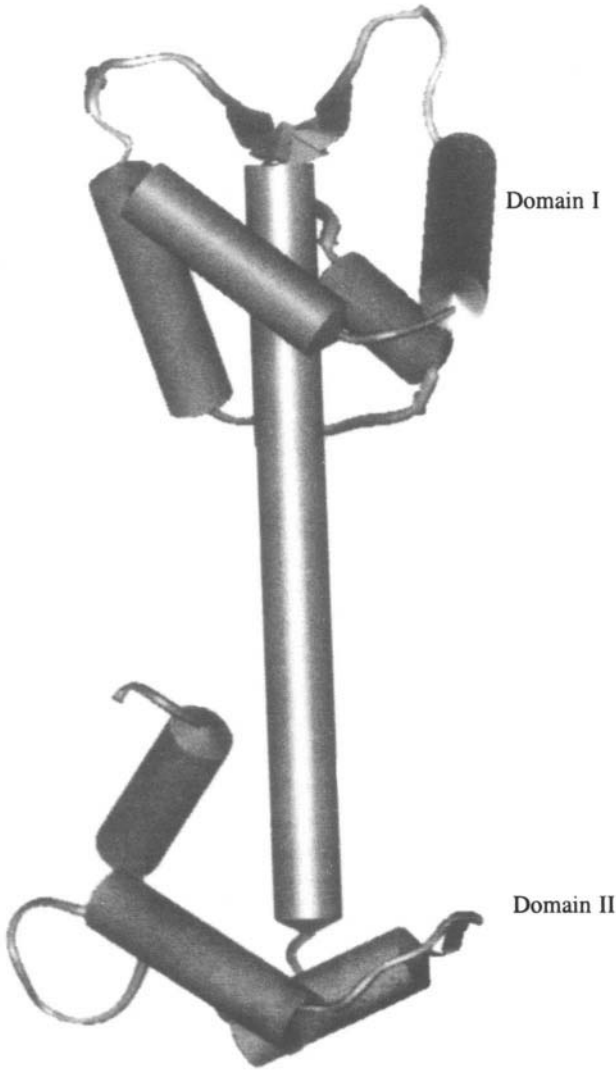


Fig. 10.38 Structure of troponin C showing the two domains

be different. For example, haemoglobin is a tetrameric protein consisting of two chains each of two different polypeptides called the **α-chain** and the **β-chain.** In most multimeric proteins, the number of subunits ranges from 2 to about 12, with a majority possessing 2 or 4 subunits. The exceptions are large enzyme complexes and viruses, which may have even 100 or more monomers. In proteins made up of only a few monomers or a small number of subunits, all the interactions that are possible by subunits must be completely fulfilled. Otherwise larger aggregates would occur. This implies that the subunits must be regularly packed around a central point. In other words most protein multimers possess point group symmetry consisting of one or more intersecting rotation axes. (Mirror inversion is also a part of point group symmetry but is not possible in protein assemblies, as it would require

the presence of d-amino acids.) Some of the rotation axes found in proteins include two-fold, three-fold, four-fold, five-fold, six-fold and eight-fold axes. A tetrahedral symmetry is also seen where the monomers are placed at the vertices of an imaginary tetrahedron (Figure 10.39). Hexameric proteins may form trigonal prisms, hexagons or octahedra. Octameric proteins may form cubes or square prisms.

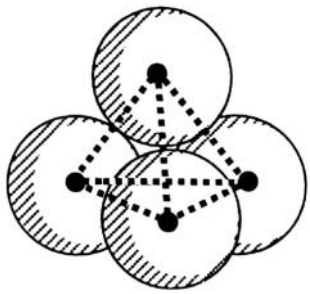


Fig. 10.39 Tetrahedral symmetry in the quaternary structure of the proteins. Each sphere represents one subunit of a tetrameric protein.

In order to attain a definite function, there are several reasons why an assembly of subunits is evolutionarily favourable, rather than giant covalently linked molecules. One of the reasons is that less DNA is needed to code such a protein, since the monomers also contain the information to self-assemble into the multimer. Again, it is easier to make smaller units in an error free fashion than large units. Similarly, defective units can be more economically discarded if they are small. A very important reason is probably that subunits with a particular function may be combined with different subunits to make up different proteins. This modular approach to building up a protein exists at the level of domains too. A particularly striking example is the structure of the nucleotide-binding region, which is almost identical in several proteins with different functions, such as glyceraldehyde-3-phosphate dehydrogenase, phophoglycerate kinase and phosphorylase.

10.3.5 Virus structure

Viruses are large macromolecular assemblies on the border between living and non-living things. The simplest virus conceptually consists of a genome (DNA or RNA) which codes for a protein that will build up a cover to protect the genome (Figure 10.40). However, the viral genome usually codes for a few other proteins too, which are necessary to allow the virus to infect its host. Even with a genome which codes for the coat protein alone, the nucleic acid molecule is too large to allow the protein to completely enclose it. If the protein were to be larger, so would the genome, requiring an even larger protein, and so on. The evolutionary answer to this problem is to construct the viral coat or 'capsid' from several copies of a single, relatively small protein. The capsid is roughly spherical in shape as for example in the common cold virus or in HIV. Other viruses are long cylinders or rod shaped, e.g. Tobacco Mosaic virus. More complex viruses such as bacteriophage *T*4 have more complex shapes.

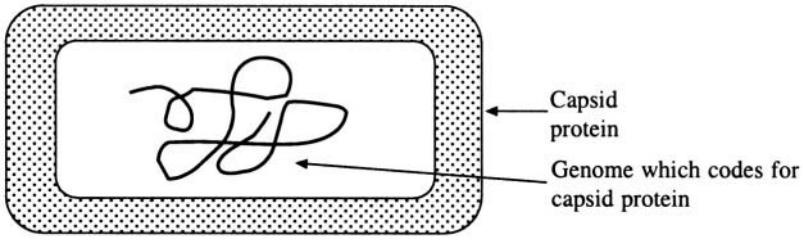


Fig. 10.40 Schematic representation of simple virus

Spherical viruses have the shape and symmetry of an icosahedron (Figure 10.41). Each of the twenty faces of the icosahedron is a triangle with a 3-fold symmetry and can accommodate 3 identical

protein molecules. The viral capsid as a whole thus has 60 subunits. If the requirement for exact symmetry is somewhat relaxed, a larger number of subunits can be accommodated. However, only certain multiples (1, 3, 4, 7) of 60 subunits are likely to occur, under the assumption that identical subunits may have quasi-equivalent environments, rather than exact equivalence. More elaborate structures are also known for spherical viruses. The protein capsid of a class of viruses known as picorna viruses (which includes polio virus and common cold virus or rhino virus) contains 60 copies of each of four different polypeptide chains.

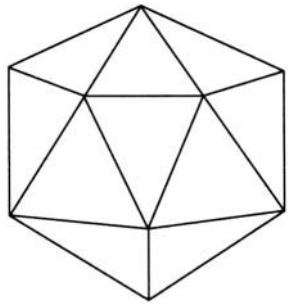


Fig. 10.41 The icosahedral shape of a 'spherical' virus

Rod-like viruses such as the tobacco mosaic virus (TMV) follow different structural principles. Here the main motif is a helix. In TMV a total of about 2130 subunits are arranged as a helix 3000 Å long and 180 Å in diameter. The pitch of the helix is 23 Å and there are 16.3 subunits per turn (i.e. 49 subunits in 3 turns) (Figure 10.42). The arrangement is reminiscent of a bunch of bananas with a long RNA molecule winding round the central stem of the bunch. Electron microscopy and electron diffraction studies indicate that the rod-like tail of bacteriophage *T*4 also consists of identical subunits arranged as a helix.

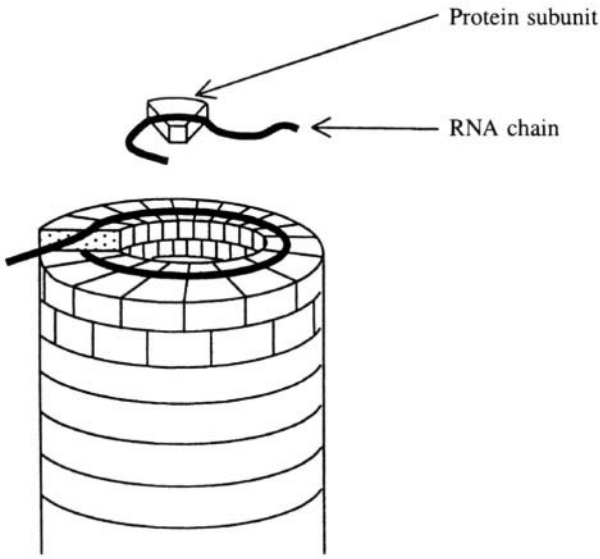


Fig. 10.42 Structure of a rod-like virus—the Tobacco Mosaic virus

11

Energy Pathways in Biology

1.1 Introduction

We studied in Chapter 1 that a process can occur spontaneously only if the sum of the entropies of the system and the surroundings increases. Therefore, for a natural process to take place, energy must be dissipated. Living systems continuously do work within themselves or on the environment. When work is being done by living system no temperature changes take place and hence these systems must spend their energy in some other form either internally or externally. The ultimate source of energy for all living systems is solar energy, which is harnessed through green plants, algae and certain types of bacteria. This process by which solar energy is converted into chemical energy that can be used by the plants and other living systems is known as photosynthesis. The various processes that are involved in the transfer of energy in the biosphere may be represented in a diagram called the energy cycle (Figure 11.1).

Living organisms do not undergo an increase in internal disorder (i.e. entropy) when they metabolise their nutrients. However this does not violate the second law of thermodynamics because the entropy of the surroundings increases. In other words living systems retain their internal order at the cost of the environment. Sometimes the entropy of a living system is said to be negative because it is rich in information content. According to information theory, information is a form of energy and is called negative entropy.

11.2 Free Energy

A continuous supply of free energy is required by living things for carrying out the active transport of ions and molecules, for muscle contraction and other mechanical work, and for the synthesis of macromolecules from their constituents. This free energy requirement is derived from the environment. Phototrops derive it from light while chemotrops obtain it in the form of foodstuff. The free energy derived from oxidation of foodstuff and from light is partly converted into a carrier molecule called

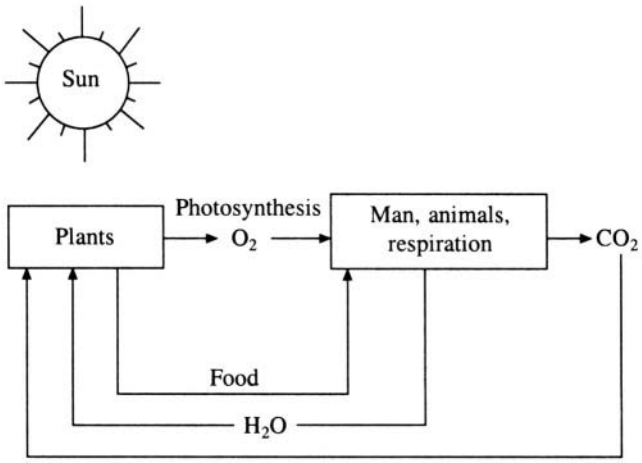


Fig. 11.1 The energy cycle

adenosine triphosphate or ATP (Figure 11.2). ATP is also known as the universal currency of free energy in biological systems. ATP consists of an adenine base, ribose sugar and a triphosphate unit. The energy rich part is the triphosphate unit. Free energy is liberated when ATP is hydrolysed to adenosine diphosphate (ADP) and orthophosphate P_i or when it becomes adenosine monophosphate (AMP) and a pyrophosphate (Figure11.3).

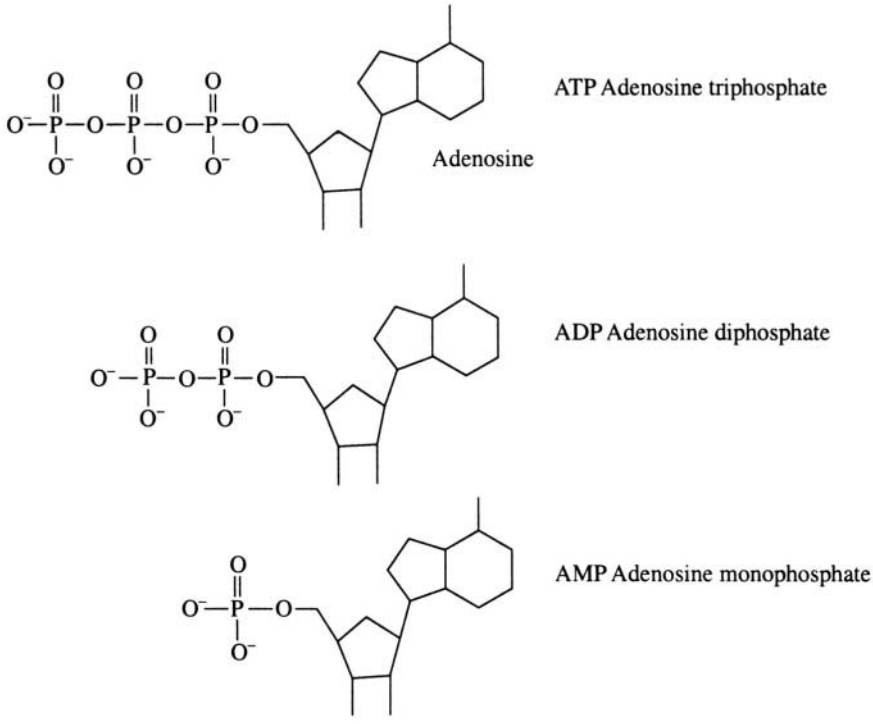


Fig. 11.2 Adenosine phosphates.

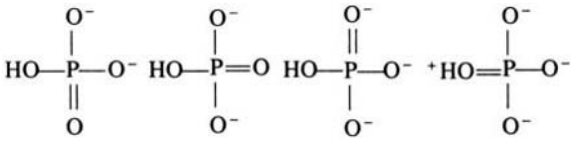


Fig. 11.3 Resonance forms of orthophosphate

$$ATP + H_2O \rightarrow ADP + P_i$$

$$ADP \rightarrow AMP + P_i$$

ATP, ADP and AMP are inter-convertible. ATP is formed from ADP when fuel molecules are oxidised in chemotrophs or when light is absorbed by phototrophs. The ATP-ADP cycle is the fundamental energy exchange in biological systems. The hydrolysis of ATP occurs as given by the following reaction

$$ATP^{4-} + H_2O \rightarrow ADP^{3-} + HPO_4^{2-} + H^+$$

At equilibrium, and under standard conditions the amount of energy released in this reaction is –7.3 K cal/mole. The function of ATP as an energy carrier molecule is dependent on the fact that ΔG^0 (change of free energy at pH 7.0 and under standard conditions) of the above reaction is negative. It is important to note that ATP is not the only phosphate compound with these features. There are others like phosphoenolpyruvate ($\Delta G^0 = -14.8$ K cal/mole), 1,3-diphosphoglycerate ($\Delta G^0 = -11.8$ K cal/mole), glucose-1-phosphate ($\Delta G^0 = -3.3$ K cal/mole), and glycerol-1-phosphate ($\Delta G^0 = -2.2$ K cal/mole). The free energy of hydrolysis of ATP falls in the middle of the range of the values for the other phosphate compounds occurring in biological systems. Hence it is an ideal candidate for energy transfer by coupled reactions with common intermediates. The hydrolysis of ATP releases free energy, which becomes useful only when it is coupled with other cellular processes that require free energy. In a typical cell, ATP molecules are consumed within a minute of their formation. Thus, ATP acts as an immediate donor of free energy rather than as an energy storage molecule. Different processes produce ATP in different biological systems. The two main processes are by (1) oxidation-reduction reaction in the inner membranes of mitochondria and (2) photosynthesis which takes place on the membranes of the grana in chloroplasts of green plants. The processes can be divided into those which depend on the oxygen availability i.e. respiration and those which are light driven i.e. photosynthesis.

11.3 Coupled Reactions

A thermodynamically unfavourable reaction can be driven by a favourable reaction, both reactions together having a net negative (and therefore favourable) difference in the free energy. These reactions are coupled by a common intermediate.
For example:

$A \Leftrightarrow B + C$	$\Delta G^0 = +6$ K cal/mole	(1)
$B \Leftrightarrow D$	$\Delta G^0 = -9$ K cal/mole	(2)
$A \Leftrightarrow C + D$	$\Delta G^0 = -3$ K cal/mole	

Under standard conditions reaction (1) can not proceed, as the ΔG^0 value is positive. But *B* can be converted to *D* as reaction (2) is thermodynamically favourable (ΔG^0 is negative). As free energy

changes are additive, the conversion of *A* to C and *D* is possible as the net ΔG^0 is negative when reactions (1) and (2) are coupled through *B,* the intermediate. The hydrolysis of ATP shifts the equilibria of coupled reactions. Therefore any thermodynamically unfavourable reaction can be converted into a favourable one by coupling it with the hydrolysis of a sufficient number of ATP molecules. For example the hypothetical reaction

$$A \Leftrightarrow B \quad (\Delta G^0 = 3 \text{ K cal/mole})$$

can be made to proceed when it is coupled to the hydrolysis reaction of ATP to ADP. The coupled reaction may be written as

$$A + \text{ATP} + H_2O \Leftrightarrow B + \text{ADP} + P_i + H^+$$

with $\Delta G^0 = -4.3$ K cal/mole. This final ΔG^0 value is obtained by adding the ΔG^0 of the $A \rightarrow B$ conversion (+ 3 K cal/mole) with that of the ATP → ADP reaction (–7.3 K cal/mole).

11.4 Group Transfer Potential

Compare the following reactions

$$\text{ATP} + H_2O \Leftrightarrow \text{ADP} + P_i + H^+ \qquad \Delta G^0 = -7.3 \text{ K cal/mole}$$

$$\text{Glucose-6-phosphate} + H_2O \Leftrightarrow \text{Glucose} + P_i \qquad \Delta G^0 = -3.3 \text{ K cal/mole}$$

The ΔG^0 for the hydrolysis of glucose-6-phosphate is much smaller than that of ATP. This means that ATP can transfer its terminal phosphoryl group to water more readily than glucose-6-phosphate. ATP is said to have a higher phosphate group transfer potential than glucose-6-phosphate. This higher potential has a structural basis. At biological pH 7:0, ATP is almost completely ionised and exists as ATP^{4-}. On hydrolysis we get

$$ATP^{4-} + H_2O \rightarrow ADP^{3-} + HPO_4^{2-} + H^+$$

Under standard conditions, ATP^{4-}, ADP^{3-} and HPO_4^{2-} will be present at concentrations of the order of 1.0 M, while the H^+ ion concentration at pH 7.0 is only 10^{-7} M. On the basis of the law of mass action, the low concentration of H^+ pulls the reaction towards the right. Further, at neutral pH there are four closely spaced negative charges in the ATP molecule. These charges repel each other and the dissociation of the terminal phosphoryl group is required, in order to release at least some of the stress. This repulsive force also prevents the products of the reaction coming together again to reform ATP. A third factor contributing to the high group transfer potential of ATP is resonance stabilisation. Orthophosphate exists in several significant resonance forms (Figure 11.3) while ATP has fewer resonance forms. In the hydrolysis of ATP the products HPO_4^{2-} and ADP^{3-} are in more stable resonance forms and contain less free energy than when they are still combined as ATP^{4-}. Thus once again the forward reaction is favoured over the reverse reaction.

In the case of glucose-6-phosphate, the reaction is

$$\text{Glucose-6-phosphate} + H_2O \rightarrow \text{Glucose} + HPO_4^{2-}$$

In this case there is no extra H^+ ion formed and chemical pressure on the reaction to proceed towards the right does not exist. Also, no repulsive force exists between glucose and HPO_4^{2-} and hence these two compounds can recombine to form glucose-6-phosphate readily. This molecule thus has a much lower transfer potential as compared to ATP.

11.5 Role of Pyridine Nucleotides

Chemotrophs get their energy from oxidation of fuel molecules like fatty acids, glucose etc. These reactions are catalysed by enzymes, which in turn require cofactors. The most common cofactors are nicotinamide adenine dinucleotide (NAD), nicotinamide adenine dinucleotide phosphate (NADP) and flavin adenine dinucleotide (FAD) (Figure 11.4). The cofactors act as carriers of electrons in the oxidation reduction reactions. In aerobic organisms electrons from the fuel molecules are transferred first to these carriers. The reduced form of these carrier molecules transfers the electrons to O through an electron transport chain located in the inner membrane of the mitochondria. ATP is formed from

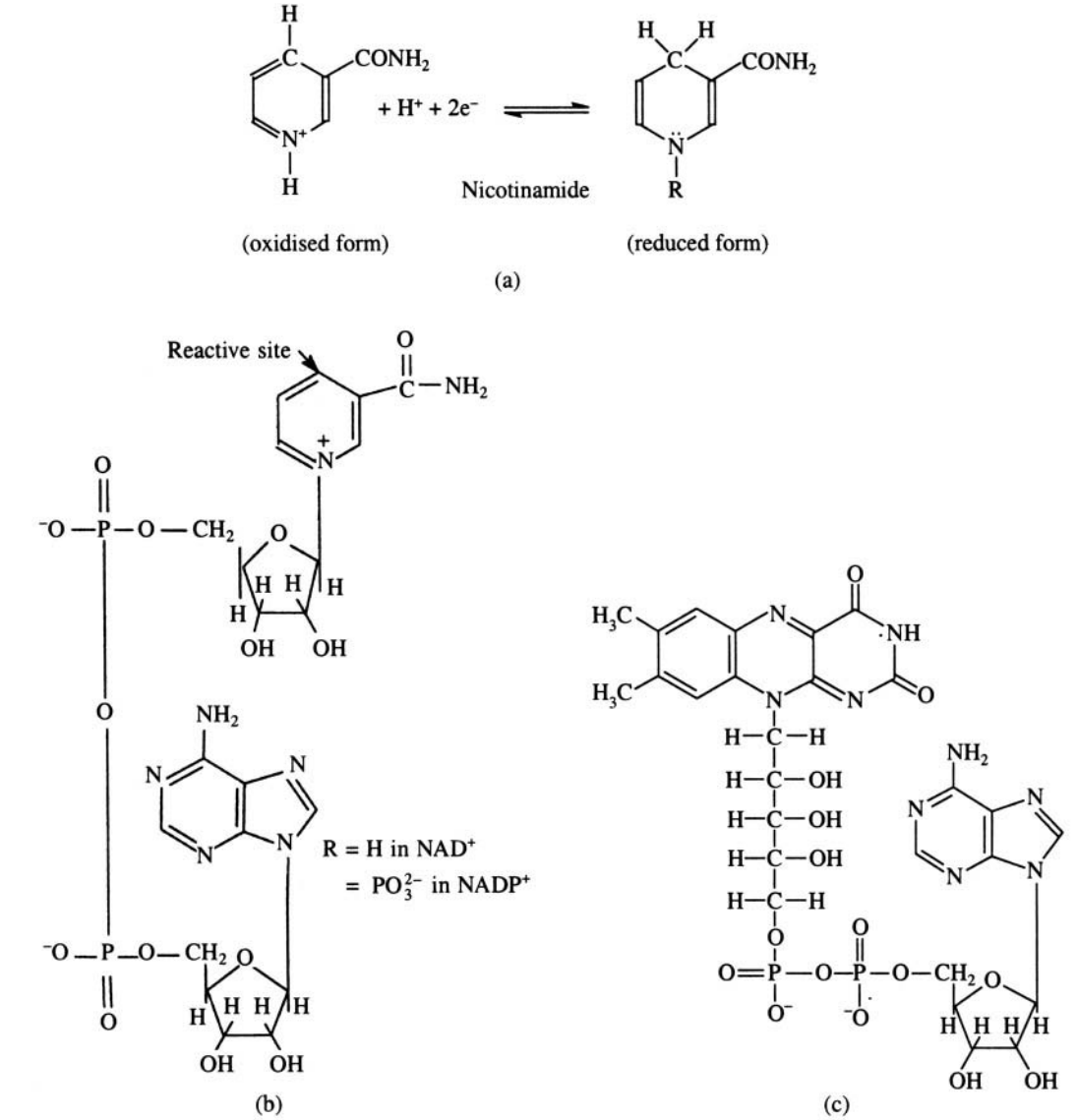
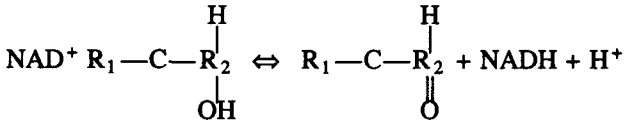
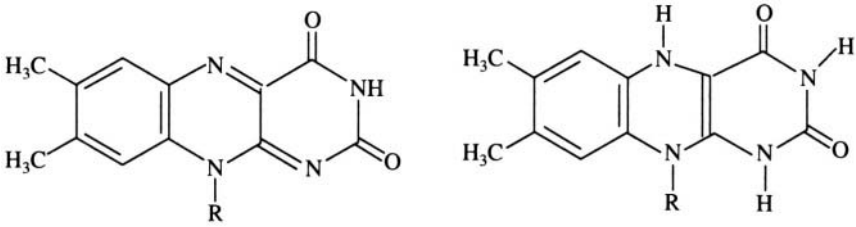


Fig. 11.4 Pyridine nucleotides. (a) Oxidised and reduced forms of nicotinamide. (b) NAD and NADP. (c) FAD

ADP and orthophosphate during this process, which is known as oxidative phosphorylation. NAD^+ accepts a hydrogen ion and two electrons from the substrate in the oxidation of the substrate molecule. In the process, NAD^+ is converted from the oxidised to the reduced state.



Similarly, the oxidised and reduced forms of FAD are shown in Figure11.5. NADP and NAD have structural similarity but their biological functions differ. NADP is used in reductive biosynthesis while NAD is used primarily in the energy metabolism. The cell has a mechanism by which NADPH can be converted to NADH and vice versa. Enzymes involved in this reaction are known as trans-hydrogenases. In the absence of catalysts, NADH, NADPH and FADH react very slowly with O_2. Similarly the hydrolysis of ATP in the absence of a catalyst is a slow process. This stability of these molecules is essential for their biological function as it provides a way for enzymes to control the reduction mechanism and hence the flow of free energy.



(a) FAD (b) $FADH_2$

Fig. 11.5 Oxidised (a) and reduced (b) forms of FAD

11.6 Photosynthesis

Photosynthesis is the process by which the energy of sunlight is captured and used to sustain life on earth. Green plants, algae and some bacteria are capable of carrying out this process which takes place in the thylakoid membranes of the organelle called chloroplast (Figurel 1.6). Light is absorbed by the pigment molecules (carotenoids, chlorophylls and phycobillins) (Figurel 1.7) in these membranes and is transferred to the so-called photosynthetic reaction centres. The light energy is converted into chemical energy at these centres. The aggregate of the reaction centre and the photosynthetic pigments is called a photosynthetic unit. Each photosynthetic unit may consist of 50 to 400 chlorophyll molecules. The absorbed light excites the pigment aggregate and this in turn transfers the energy to the reaction centre. The photosynthetic reaction takes place in two phases (Figure 11.8). In the first phase a reductant is produced in the presence of light. In the second stage this reductant is utilised to form sugar in the absence of light. The excitation energy transfer from chloroplast to the next member in the chlorophyll cluster (light harvesting or antenna molecules) takes place through inductive resonance transfer till it reaches a specialised chlorophyll molecule at the reaction centre. The reaction centre is a protein-chlorophyll complex system. These reactions are described by the equations:

$$S + h\nu_a \rightarrow S^*$$

$$S^* + A \rightarrow S + A^*$$

$$A^* \rightarrow A + h\nu_e$$

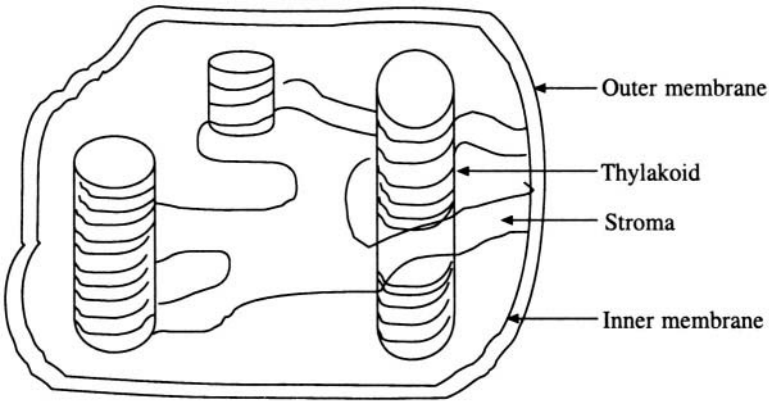
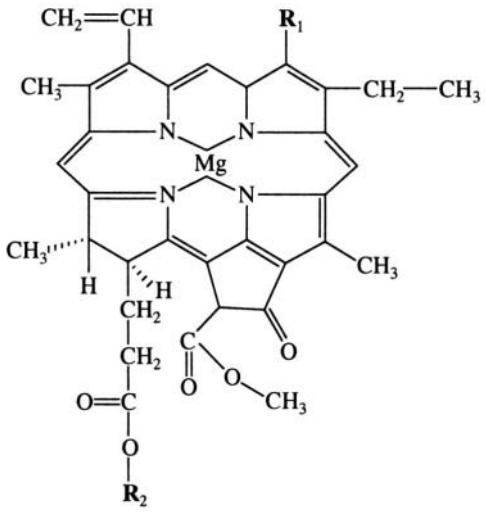


Fig. 11.6 A schematic sketch of a chloroplast showing some of the structures.



R_1 = –CH_3 chlorophyll a
= –CHO chlorophyll b

$$R_2 = -CH_2-CH=\underset{\displaystyle CH_3}{\overset{|}{C}}-CH_2-(CH_2-CH_2-\underset{\displaystyle CH_3}{\overset{|}{CH}}-CH_2)_2-CH_2-CH_2-CH\begin{matrix}CH_3\\ \diagup\\ \diagdown\\ CH_3\end{matrix}$$

Fig. 11.7 Chemical structures of the pigment molecules in chloroplasts—chlorophyll a and b.

where $h\nu_a$ is the absorbed light, $h\nu_e$ the emitted light, S is the sensitiser molecule, i.e. the reductant, and A is the acceptor molecule, and S^* and A^* are the excited stages. The photosynthetic reaction can be written as

$$nH_2O + nCO_2 \xrightarrow{\text{Light}} [CH_2O]_n + nO_2$$

However, photosynthetic bacteria do not produce oxygen but they give off elemental sulphur. The two phases of photosynthetic reaction described above are known as light reactions and dark reactions.

Two types of reactions involving light take place during photosynthesis and are called photosystem I and photosystem II respectively. In photosystem I, NADPH is produced. In photosystem II, O_2 is evolved and a reductant is generated. Photosystem I has a greater number of a particular type of chlorophyll molecule called chlorophyll *a* and is maximally excited at longer wavelengths. Photosystem II is maximally activated at wavelengths shorter than 680 nm and has more chlorophyll *b*. All photosynthetic cells of higher plants and cyanobacteria that evolve oxygen contain both photosystems I and II. Those bacteria that do not evolve the oxygen contain only photosystem I. Since photosystem I is maximally excited by light of wavelengths 700 nm, the reaction centre of this photosystem is called as P700. Similarly, the reaction centre of the photosystem II is called P680. Figure 11.9 shows the interactions between and within the two photosystems.

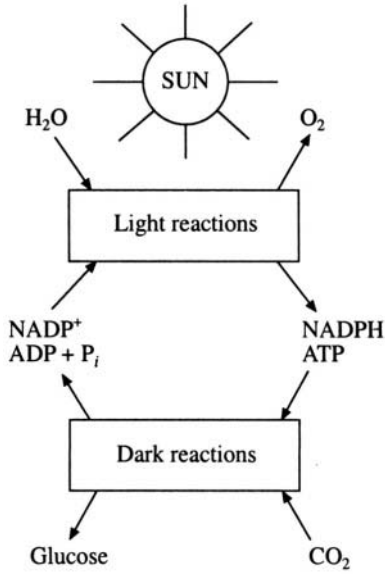


Fig. 11.8 Two stages of photosynthetic reaction

When a photon is absorbed by one of the chlorophyll molecules, the molecule goes from the ground state to an excited state. Generally when the molecule returns to the ground state it emits radiation known as fluorescence. There exists a coupled event wherein the de-excitation of one chlorophyll molecule is accompanied by the excitation of another molecule. This process is also

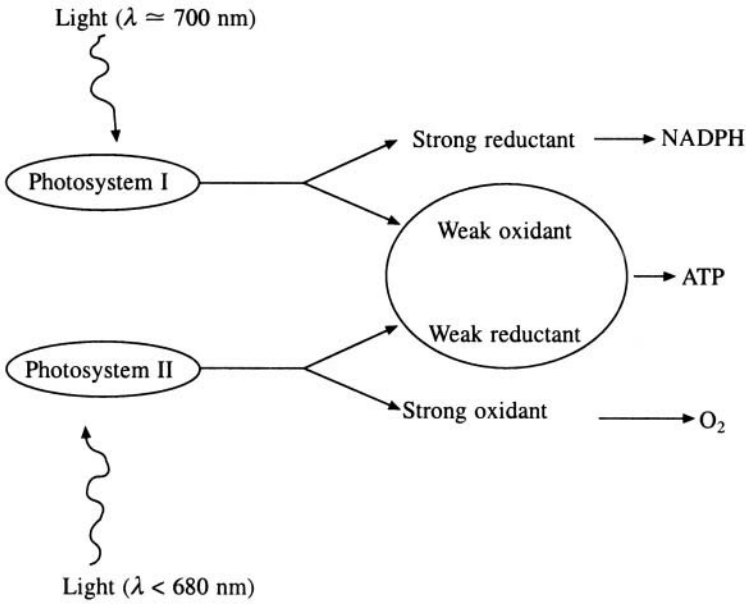


Fig. 11.9 Interactions between the two photosystems

known as sensitised fluorescence. As shown in Figure 11.10, the energy levels of the excited state of the reaction centre are lower than that of other chlorophylls. Hence it acts as an energy trap. This transfer of energy takes place in less than 10^{-10} sec.

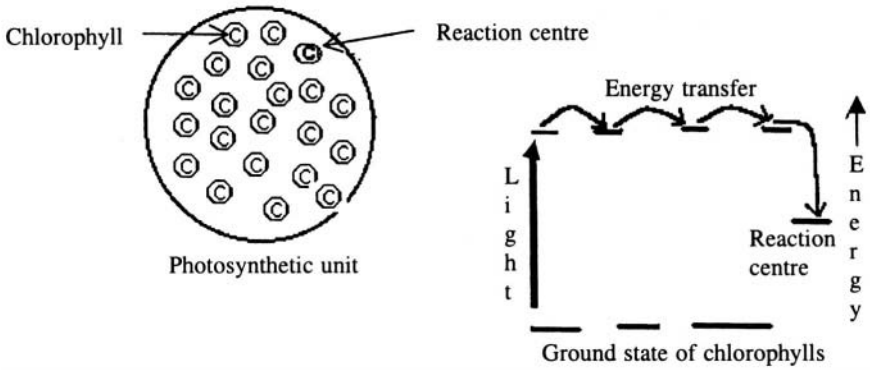


Fig. 11.10 Photosynthetic unit and energy level diagram of the antenna chlorophylls.

Recently, by the combined effort of X-ray crystallographers, spectroscopists and molecular geneticists, the mechanism of photosynthesis in the bacteria of Rhodopseudomonas family has been elucidated (Figure 11.11). Photosynthesis in these bacteria differs from that of higher plants in that no oxygen is given off during the process. However, the results obtained can be extrapolated to higher plants due to many other similarities. The study throws light on the mechanism of electron transport from one end of the reaction centre to the other end. The photosynthetic reaction centre primarily consists of a protein, embedded in the membrane of the photosynthetic vesicles. One end of the reaction centre is near the inner surface of the membrane and the other end is near the outer surface. Several small molecules known as prosthetic or helper molecules are also partially embedded in the protein complex. Photoelectrons flow through the prosthetic molecules during photosynthesis. The photosynthetic reaction centre has a two-fold symmetry axis running from one surface of the membrane to the other. The prosthetic groups are arranged in two spirals on either side of the central protein. When light falls on a special pair of chlorophyll molecules situated at the end of the reaction centre near the inner surface, an electron within the pair absorbs the energy and moves to the next prosthetic molecule viz. pheophytin. This step takes 2 picoseconds to complete. During this process the electron passes through another chlorophyll molecule but does not get attached to it. As an electron has left the special pair of chlorophyll molecules, they now have acquired a positive charge. Next the electron moves from pheophytin to a quinone molecule (Q_A) which is at the end of the spiral. The time taken for this process is about 4 picoseconds. Lastly, the electron moves through the central core of the protein to reach the quinone molecule of the other spiral (Q_B) through which electrons do not flow. This step is comparatively very slow and takes about 100 microseconds. In the meantime a cytochrome molecule donates an electron to the reaction centre and neutralises it. Thus the reaction centre is ready to receive another photon and the process repeats itself. When the quinone molecule has acquired two electrons it detaches itself from the reaction centre and participates in the later stages of the photosynthesis, which takes place on the outer surface of the membrane. Thus there will be two cytochrome molecules with positive charges at the inner surface and two electrons near the outer surface. The separation of the two charges represents stored energy. Later, the charge separation drives chemical reactions that are coupled to the bacterial metabolism. In most cases, the transfer of

electrons can take place in both the directions and then the energy is lost as fluorescent radiation. But in the photosynthetic reaction centre the forward reaction is more than eight orders of magnitude faster than the backward reaction and hence the electric charges induced by light energy remain separated. The efficiency of the reaction is as high as 50% i.e. the energy stored in the separated charges is about half of the energy of the incident photons.

11.6.1 Photosystem I

As mentioned earlier, the reaction centre of photosystem I (P700) consists of chlorophyll *a* molecules. When P700 is excited by the absorption of light by the chlorophyll molecules, it transfers an electron to bound ferrodoxin, which is a Fe_4S_4 type iron-sulphur protein. The oxidation-reduction potential of

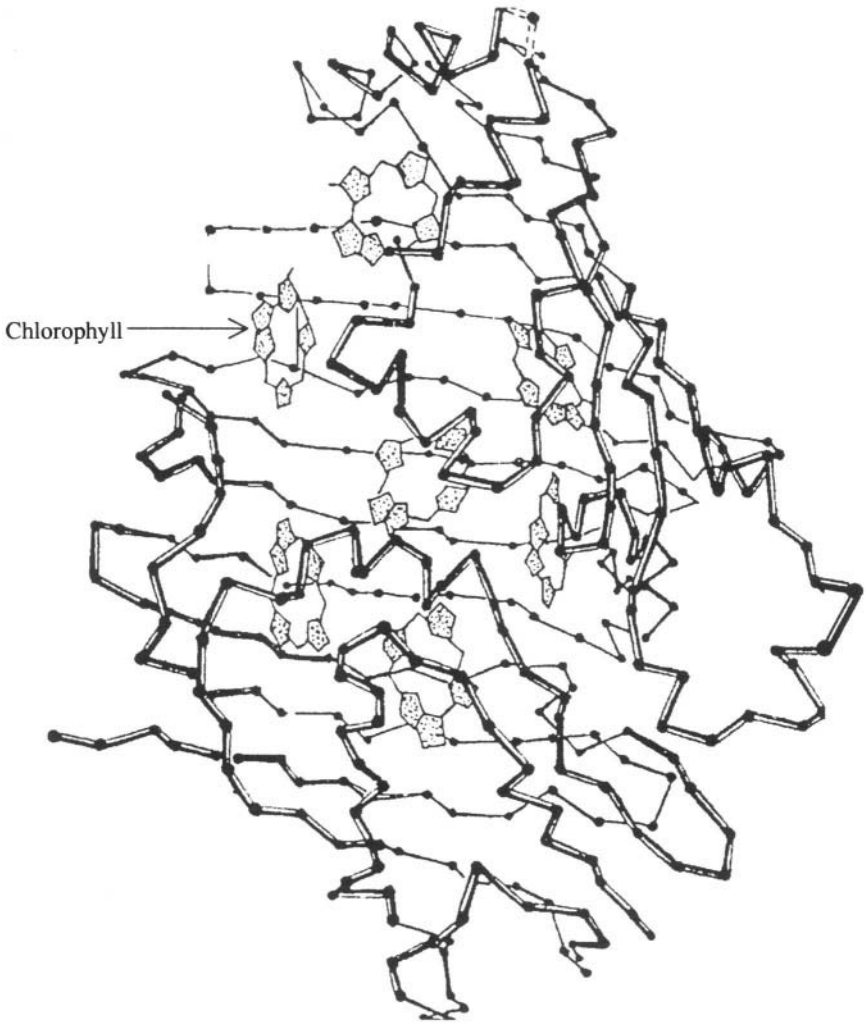


Fig. 11.11 (a) Structure of a bacteriochlorophyll-protein complex.

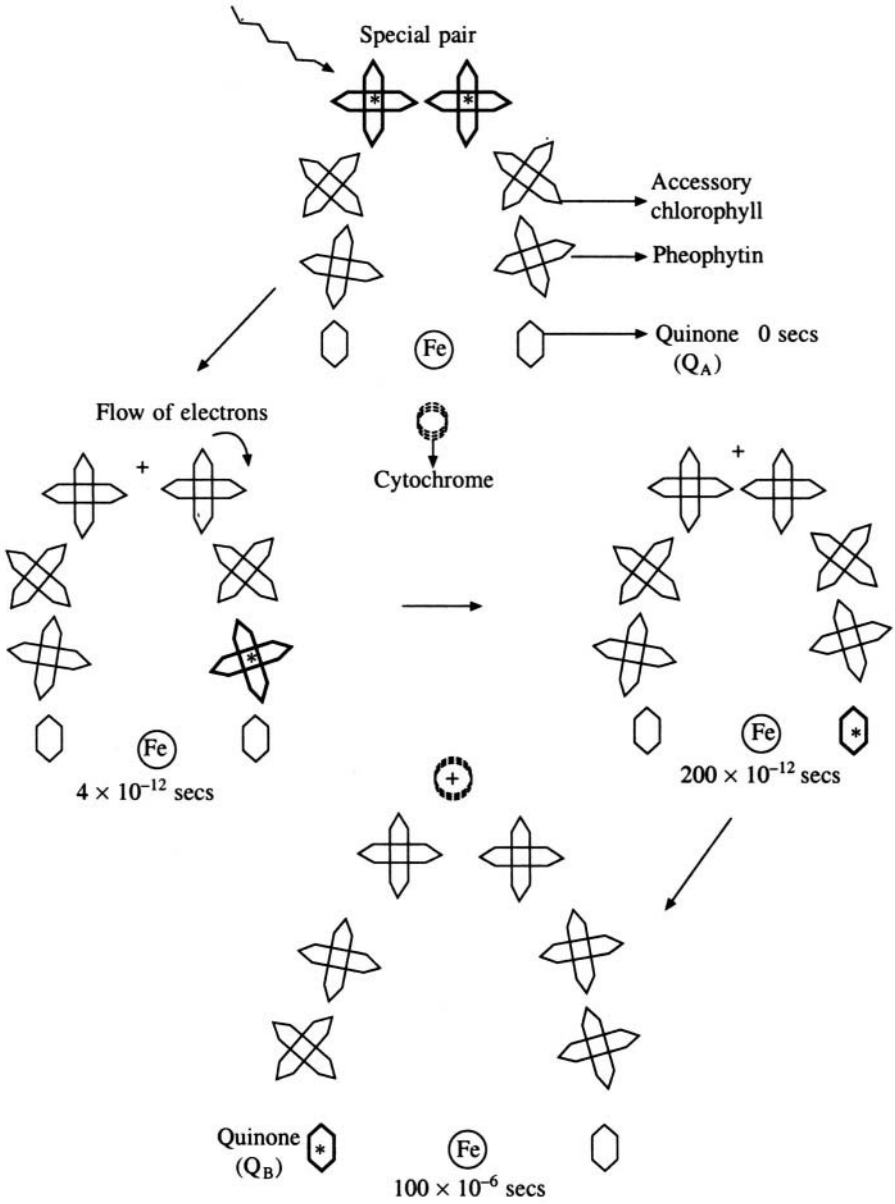
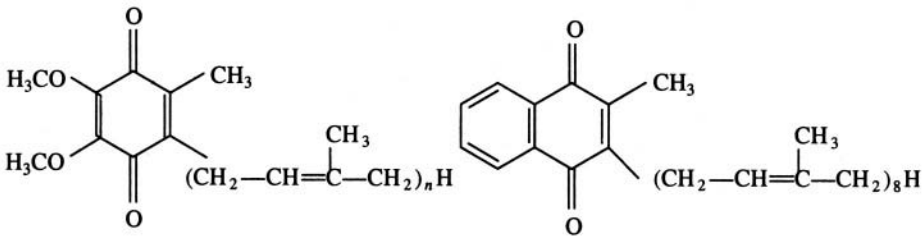


Fig. 11.11 (b) Mechanism of electron transport at photosynthetic reaction centre. Excited groups are shown in bold with a star.

P700 changes from +0.4 V to –0.6 V when it is excited by light. From the bound ferrodoxin the electron moves to soluble ferrodoxin and then to **NADP⁺** reducing it to NADPH. During this process the iron atom of ferrodoxin is alternatively oxidised and reduced. The reduction of **NADP⁺** to NADPH requires two electrons whereas ferrodoxin carries only one electron at a time. Hence electrons from two reduced ferrodoxins come together to convert **NADP⁺** to NADPH (Figure 11.12).



Ubiquinone (Q_A) Menaquinone (Q_B)

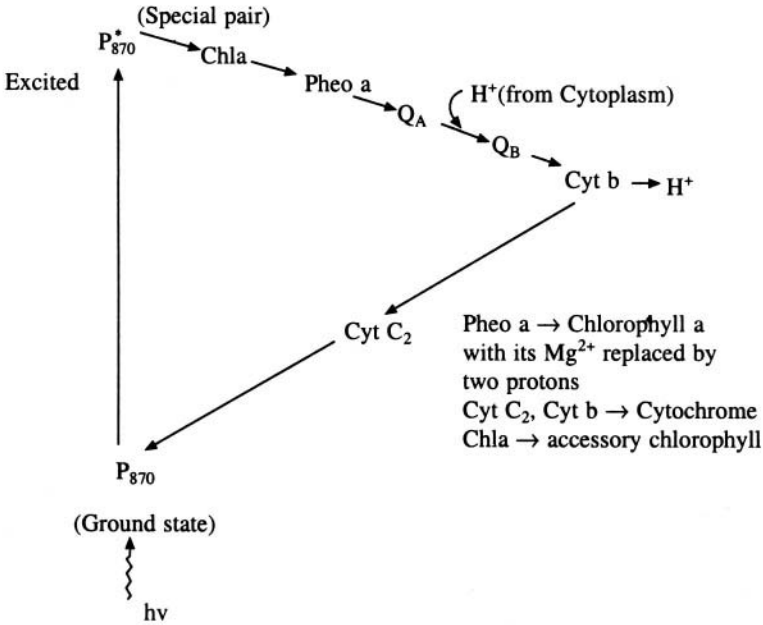


Fig. 11.11 (c) Photosynthetic reactions

11.6.2 Photosystem II

The photosystem II (P680) is not very well understood excepting for the fact that the redox potential of this reaction centre is +0.8 V. Photosystem II produces a strong oxidant (A^+) and a weak reductant (B^-) and transfers an electron to P700. A^+ attracts electrons from water and gets neutralised. In this process O is released.

$$4A^+ + 2H_2O \rightarrow 4A + 4H^+ + O$$

The electron from photosystem II goes through a bound molecule of plastoquinone, then through mobile plastoquinones, through cytochromes b_{559} and c_{552}, finally to plastocyanin, which transfers it to photosystem I. Thus photosystem I, which was in an oxidised state after absorption of light, is now reduced and is ready for the next cycle of photo reaction. A proton gradient is developed across the thylakoid membrane when electrons flow through the above carriers and, as described earlier, this gradient drives the formation of ATP. The electrons from P700 can also follow a cyclic path, if there is insufficient $NADP^+$ to accept electrons from reduced ferrodoxin. In the cyclic path the electron gets

transferred through cytochrome b_{563}, c_{552}, and plastocyanin, back to P700. Thus NADPH and ATP produced during light reactions are now available for subsequent dark reactions in which CO_2 is converted to carbohydrate.

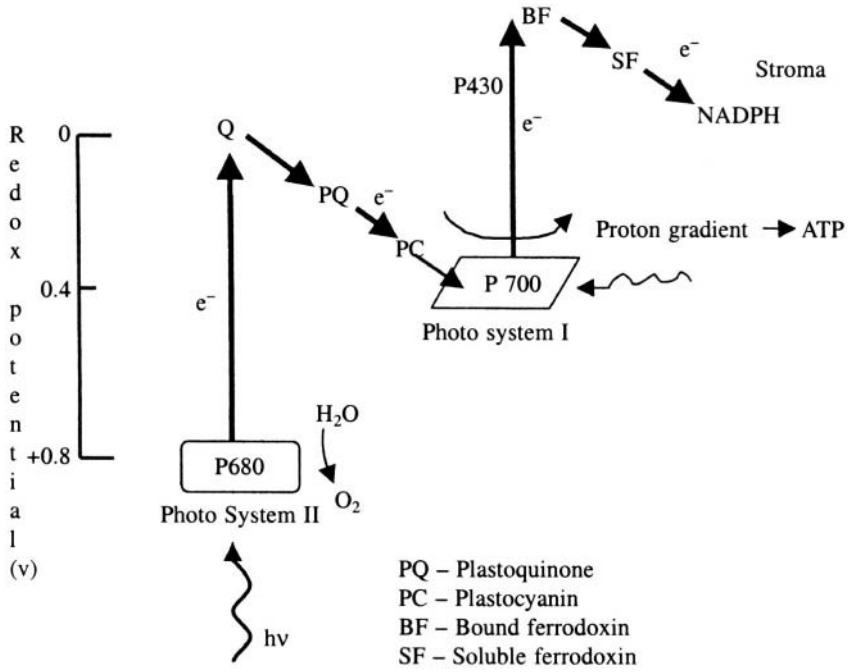


Fig. 11.12 Electron flow in photophosphorylation

11.6.3 Photophosphorylation and carbon fixation

In the second phase of photosynthesis, which occurs in the absence of light, CO_2 is fixed into carbohydrate through a series of reactions known as the Calvin cycle. CO_2 condenses with a five-carbon compound (ribulose-1,5-diphosphate) to form an unstable six-carbon compound which hydrolyses to two three-carbon compounds (3-phosphoglycerate). The reaction can be written as

$$C_5 \rightarrow C_6 \rightarrow C_3 + C_3$$

The cycle is shown in Figure 11.13. Three ATP molecules and two NADPH molecules are consumed in the process of CO_2 fixation as a sugar. As only one carbon is reduced in each cycle, a total of six cycles will be required to form hexose. The complete reaction is given by the equation

$$6CO_2 + 18ATP + 12NADPH + 12H_2O \rightarrow C_6H_{12}O_6 + 18ADP + 18P_i + 12NADP^+ + 6H^+$$

To calculate the efficiency of the system, consider the following: ΔG^0 for CO_2 reduction is 114 K cal/mole. $NADP^+$ reduction to NADPH in photosytem I requires two electrons and four photons. Photosystem II replenishes the lost electrons of photosystem I. It is known from a study of the photosystems that, for one electron to be transferred from H_2O to $NADP^+$, two photons must be absorbed by each of the two photosystems. Thus a total of eight photons will have to be absorbed for producing the required number (2) of NADPH molecules, per CO_2 fixed. The energy of the photons may vary from 72 K cal/mole at 400 nm to 41 K cal/mole at 700 nm. If the energy of the photon at 600 nm (47.6 K cal/mole)

is taken as the average, the total energy absorbed from the 8 photons is 8 x 47.6 = 381 K cal/mole. The efficiency of the photosynthetic reaction at 600 nm wavelength is 114/381 = 30% under standard conditions.

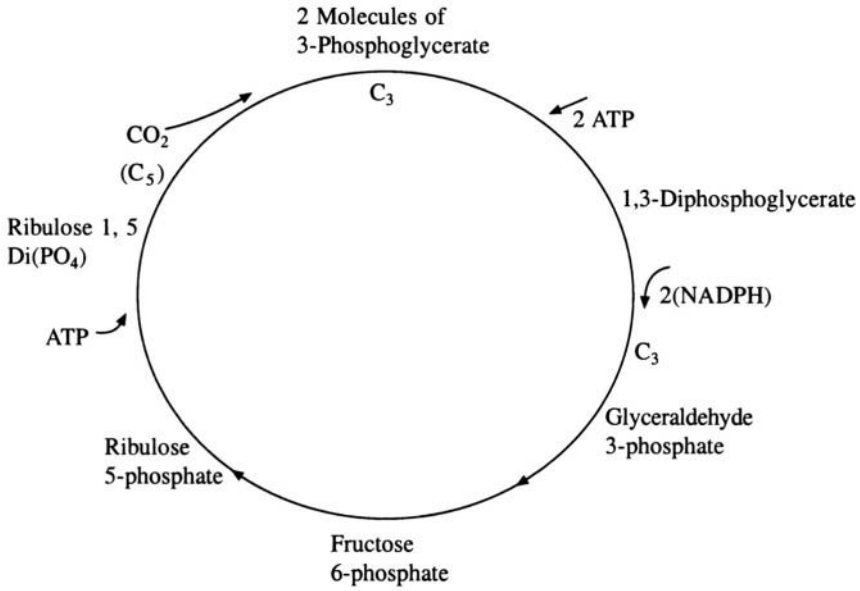


Fig. 11.13 Calvin cycle.

11.7 Energy Conversion Pathways

ATP and NADPH are continuously generated and degraded. These processes are regulated and the biosynthesis and the degradation have well-separated distinct pathways. When a high level of ATP is available then anaerobic reactions, which utilise ATP, are stimulated and ATP generating pathways are inhibited. The reverse occurs when the level of ATP is low.

11.7.1 Oxidation

Oxidation of food is always coupled with the formation of ATP and this process can take place even in the absence of oxygen. In some organisms anaerobic oxidation is the only available pathway for energy conversion whereas in others both anaerobic as well aerobic energy conversion can take place. (Figure 11.14)

11.7.2 Glycolysis

Glycolysis takes place in the cytosol (Figure 11.15). Glycolysis is a metabolic pathway in which glucose is converted into pyruvate with the concomitant production of ATP. This mechanism precedes a sequence of other oxidation reactions in higher organisms. Figure 11.16 gives in detail the role of NAD in the glycolytic pathways in anaerobic and aerobic metabolism respectively. When ethanol or lactate is produced from glucose in anaerobic organisms it is known as fermentation. The glycolytic pathway has a dual role in that it degrades glucose to form ATP and also provides building blocks for reactions like formation of long chain fatty acids. Glycolysis is controlled by essentially irreversible reactions, such as those catalysed by hexokinase, phosphofructokinase and pyruvate kinase. In these reactions, two molecules of ATP are generated during the conversion of glucose to pyruvate.

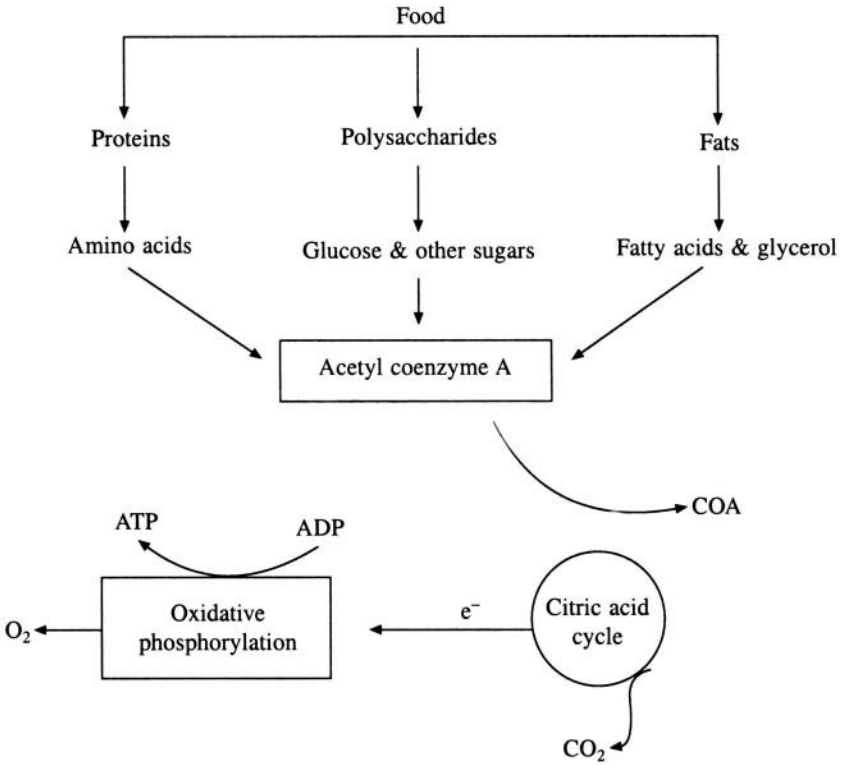


Fig. 11.14 Generation of energy from food

In aerobic organisms, glycolysis is followed by respiration, which takes place in the mitochondria (Figure 11.17). Mitochondria have two membranes. The outer membrane is permeable to small molecules. The inner membrane forms several inner folds thus increasing the total surface available. These folds are called cristae. They enclose an inner compartment known as the matrix. There are several enzymes in the membrane. The respiratory reactions take place in two steps, the first of which takes place in the matrix. This cycle is named after its discoverer Hans Krebs as Krebs cycle. The second step is part of the respiratory cycle and takes place in the cristae.

11.7.3 The Krebs cycle

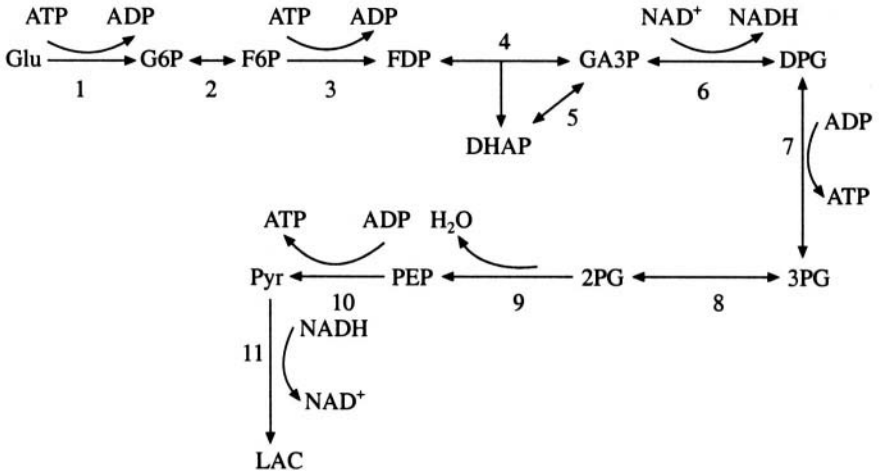
The Krebs cycle is also known as the citric acid cycle, or the tricarboxylic acid cycle (TCA). In addition to the oxidation-reduction reactions that lead to the generation of ATP, the citric acid cycle provides several intermediates, which are branch points for the synthesis of various compounds required by the cell, like amino acids. Most fuel molecules enter this cycle as acetyl coenzyme *A* (acetyl CoA). The reaction for the formation of acetyl CoA is as follows.

$$\text{Pyruvate} + \text{CoA} + \text{NAD}^+ \rightarrow \text{Acetyl CoA} + \text{CO}_2 + \text{NADH}$$

The Krebs cycle can be written as

$$\text{Acetyl CoA} + 3\text{NAD}^+ + \text{FAD} + \text{GDP} + P_i + 2\text{HO} \rightarrow 2\text{CO}_2 + 3\text{NADH} + \text{FADH}_2 + \text{GTP} + 2\text{H}^+ + \text{CoA}$$

Two carbon atoms from acetyl CoA enter the cycle and condense with oxaloacetate (a four-carbon



Enzymes	Substrates
1. Hexokinase	Glu – Glucose
2. Phosphoglucose isomerase	G6P – Glucose-6-phosphate
3. Phosphofructokinase	F6P – Fructose-6-phosphate
4. Aldolase	FDP – Fructose 1, 6-diphosphate
5. Triose phosphate isomerase	GA3P – Glyceraldehyde-3-phosphate
6. Glyceraldehyde-3-phosphate dehydrogenase	DHAP – Dihydroxy acetone phosphate
7. Phosphoglyceratekinase	DPG – 1, 3 Diphosphoglycerate
8. Phosphoglyceromutase	3PG – 3-Phosphoglycerate
9. Enolase	2PG – 2-Phosphoglycerate
10. Pyruvate phosphokinase	PEP – Phospho enol pyruvate
11. Lactate dehydrogenase	Pyr – Pyruvate
	Lac – Lactic acid

Fig. 11.15 Glycolysis

molecule) to form the six-carbon molecule citrate. Citrate is isomerised to isocitrate and oxidative decarboxylation of isocitrate gives **α-ketoglutarate** (a five-carbon molecule). One molecule of CO_2 is evolved during this reaction. A second molecule of CO_2 is given out when **α-ketogluterate** is oxidatively decarboxylated to succinyl CoA (again a four-carbon molecule). In the next stage of the reaction GTP is generated and succinate is formed. Succinate is then oxidised to fumarate (a four-carbon molecule) which is subsequently hydrated to malate. Finally oxaloacetate is regenerated from malate by oxidation. Figure 11.18 gives a schematic sketch of this cycle. In all there are four oxidation-reduction reactions and 3 pairs of electrons are transferred to NAD^+ and one pair to FAD. These reduced electron carriers are oxidised in the electron transport chain to form eleven molecules of ATP. One high-energy phosphate molecule, viz. GTP is produced during the cycle. Thus twelve high-energy phosphate molecules are produced in the oxidation of acetyl group into H_2O and CO_2. When NADH and FADH lose their electrons in the electron transport chain, NAD^+ and FAD are regenerated. As already mentioned intermediates for the biosynthesis of amino acids, fatty acids etc., are also formed during the Krebs cycle. The loss of intermediates of the cycle due to biosynthesis is constantly compensated, since many of the branching off reactions are reversible. In addition to this there are reactions known as anaplerotic (fill up) reactions. For example, if oxaloacetate is converted into amino acids then the supply of oxaloacetate for the Krebs cycle is replenished by the carboxylation of pyruvate.

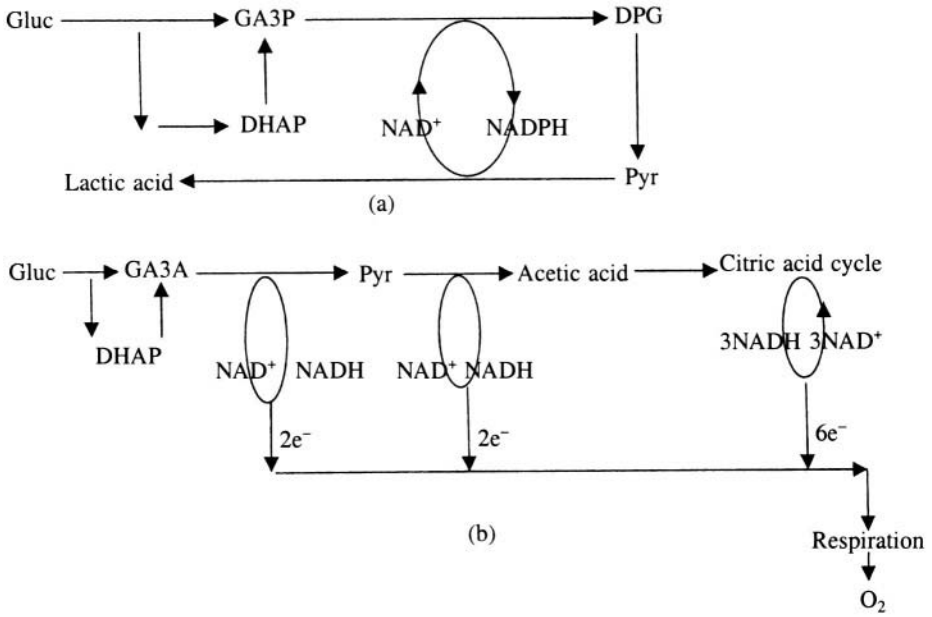


Fig. 11.16 Role of NAD in (a) anaerobic metabolism and (b) aerobic metabolism.

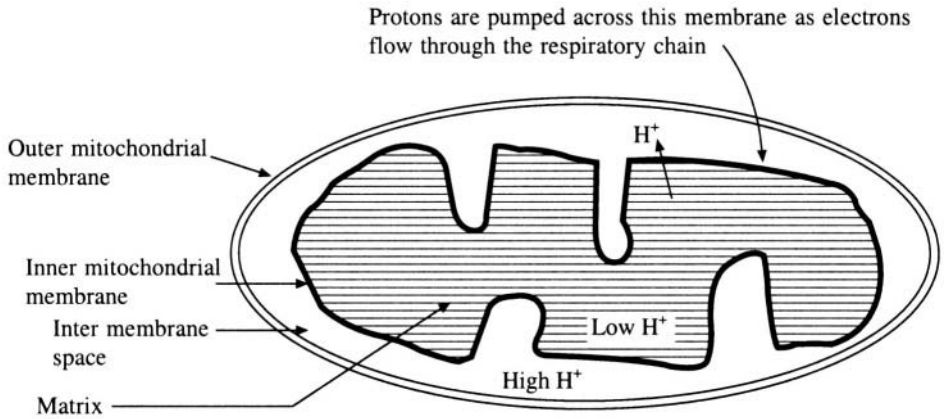


Fig. 11.17 Schematic diagram of mitochondria.

$$\text{Pyruvate} + CO_2 + \text{ATP} \rightarrow \text{Oxaloacetate} + \text{ADP} + P_i + 2H^+$$

The biosynthetic intermediates in the Kerb's cycle are shown in Figure 11.19.

11.7.4 The respiratory chain

The respiratory chain consists of a series of oxidation-reduction enzymes bound to the cristae membrane of mitochondria. The NADH and $\mathbf{FADH_2}$ molecules formed in glycolysis, the citric acid cycle, etc. are energy rich molecules because each is capable of transferring two electrons to molecular oxygen, thus liberating a large amount of energy. This energy can be utilised to generate ATP. This is the major source of ATP molecules in aerobic organisms. The oxidation of NADH, $\mathbf{FADH_2}$ and phosphorylation are coupled processes. Figure 11.20 shows the respiratory cycle. The proteins involved in the chain are

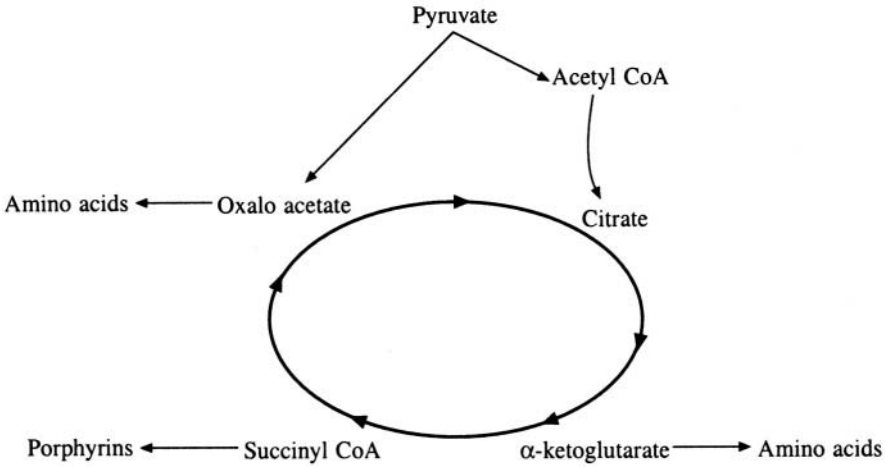
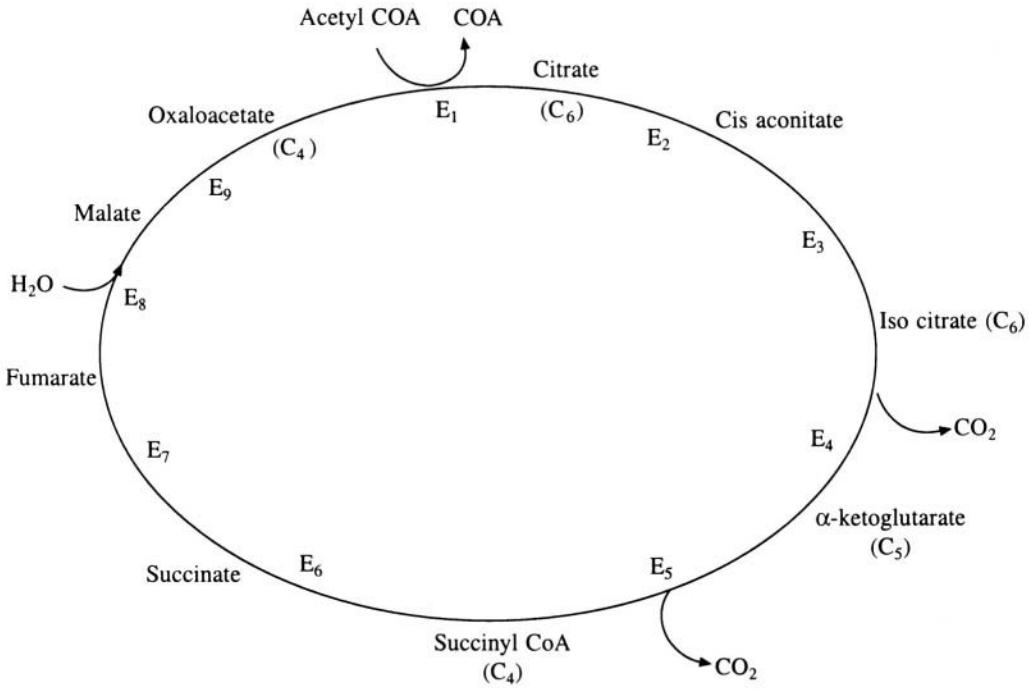


Fig. 11.18 Biosynthetic roles of Kreb's cycle



Enzymes (E)

1. Citrate synthase
2. Aconitase
3. Isocitrate dehydrogenase
4. Oxalo succinate decarboxylases
5. α-ketoglutarate dehydrogenase
6. Succinyl-COA synthase
7. Succinate dehydrogenase
8. Fumarase
9. Malate dehydrogenase

Fig. 11.19 Biosynthetic intermediates in the Kreb's cycle

FMN or NAD dehydrogenases, FAD or succinate dehydrogenases, cytochromes and iron-sulphur protein complexes. Electron transfer from NADH to O_2 takes place through a number of electron carriers like flavins, iron-sulphur complexes, quinones and hemes. It has been observed that when the mitochondrial (cristae) membrane is treated with specific detergents, the components separate into four structured complexes. Complex I contains NADH dehydrogenase and four Fe–S centres. Complex II consists of succinate dehydrogenase and its Fe–S centres. Complex III is made of cytochromes *b* and *c,* and a specific Fe–S centre. Cytochrome *a* and a_3 constitute complex IV. Ubiquinone is the connecting link between complex I, II and III while cytochrome *c* is the link between complex III and IV.

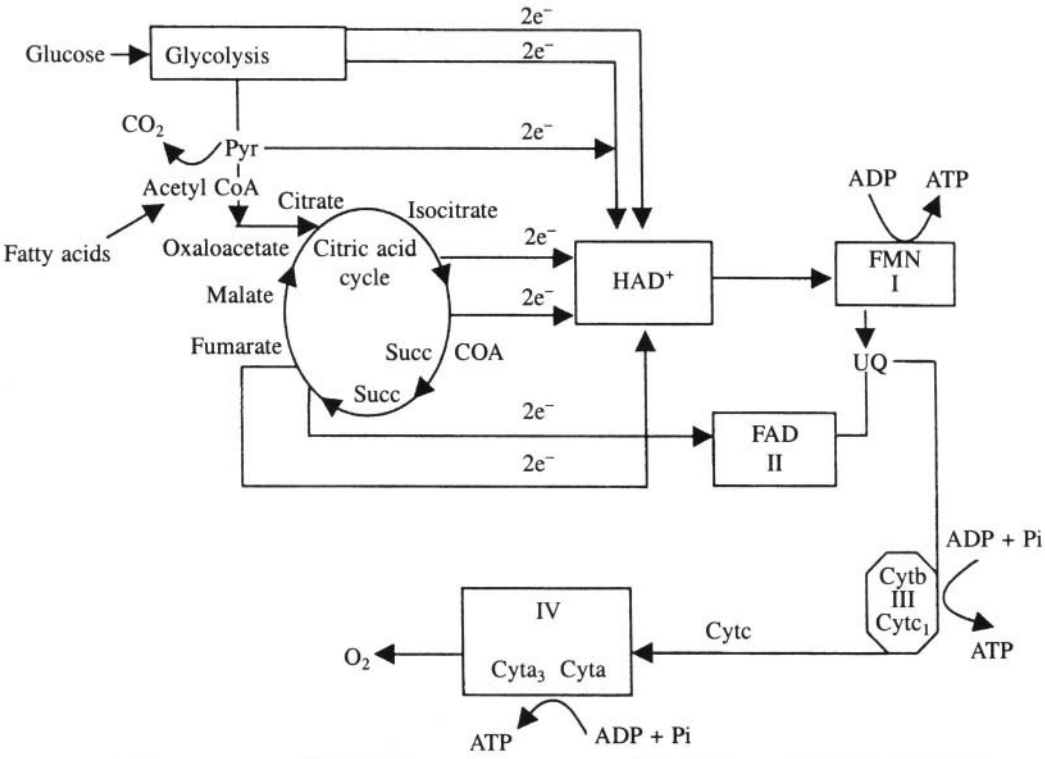


Fig. 11.20 Respiratory cycle

As discussed in Chapter *I*, chemical reactions in which electrons are transferred from one molecule to the other are termed as oxidation-reduction reactions. A redox (oxidation-reduction) reaction can be written as:

$$\text{Electron donor} \Leftrightarrow e^- + \text{Electron acceptor}$$

e.g.

$$Fe^{2+} \Leftrightarrow e^- + Fe^{3+}$$

Fe^{2+} and Fe^{3+} together form a conjugate redox pair. The standard redox potential (E_0') is defined

as the electromotive force (emf) in volts measured by an electrode placed in a solution containing both electron donor and its conjugate acceptor in 1.0 M concentration at 25°C and pH = 7.0. Increasingly negative redox potential means an increasing tendency to loose electrons. Similarly increasingly positive potential indicates an increasing tendency to accept electrons. Thus electrons will tend to flow from a relatively electronegative conjugate pair like **NADH/NAD$^+$ (E_0' = – 0.32 V)** to more electropositive redox pair like reduced cytochrome c/oxidised cytochrome ***c*** **(E_0' = 0.23V)** (Figure. 11.21). This electron flow leads to a loss of free energy, which is utilised for conversion of **ADP → ATP.** The complete equation for oxidative phosphorylation can be written as

$$NADH + H^+ + 2O_2 + 3Pi + 3ADP \rightarrow NAD^+ + 3ATP + 4H_2O$$

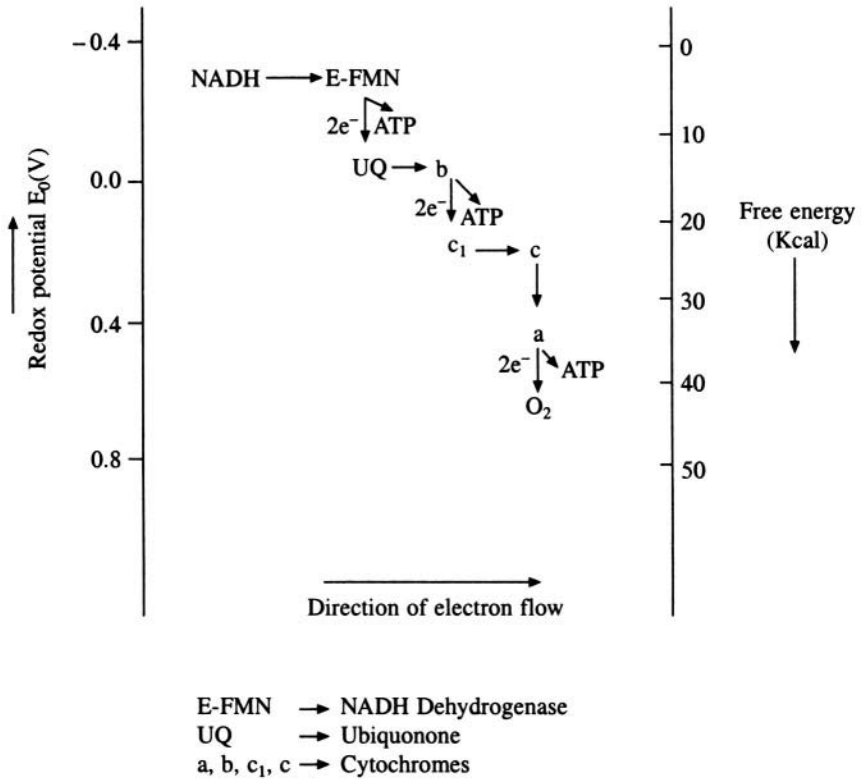


Fig. 11.21 Direction of electron flow in the respiratory chain of mitochondria

ATP formation by oxidative phosphorylation is regulated by the **[ATP]/[ADP][P_i]** ratio. This ratio is known as the mass action ratio of the ATP system. The square brackets denote molar concentrations. Using absorption spectrophotometry one can follow oxidation-reduction reactions of the electron carriers. For example cytochromes, which absorb light in the visible region, undergo spectral changes when the enzyme changes its redox state, and these can be followed. Such coupled reactions can be inhibited by compounds like 2,4-dinitrophenol (DNP), certain classes of antibiotics etc. A protein called coupling factor is required for the coupled reaction to proceed. These coupling factors are located inside the matrix space and connected to the inner membrane by a narrow stalk (Figure 11.22). The coupling factor, which is referred to as ***F_1***, catalyses the synthesis of ATP and is also known as ATP synthetase complex. Electron flow in the electron transport chain results in the

pumping of protons across the inner mitochondrial membrane, i.e. from matrix to the cytoplasm (Figure 11.17). Thus the H^+ ion concentration is greater in the cytoplasm and a potential is generated across the membrane, with the cytoplasmic side positive. This proton motive force drives the synthesis of ATP by ATP synthetase complex. Oxidative phosphorylation generates 32 out of the 36 molecules of ATP produced when glucose is completely oxidised to CO_2 and H_2O. The electron flow from $NADH \rightarrow CO_2$ is an exergonic process.

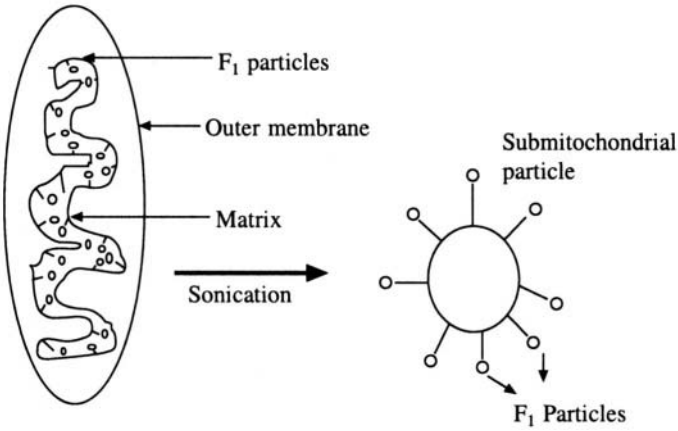


Fig. 11.22 Cross section of a mitochondrion

11.8 Membrane Transport

The inner mitochondrial membrane is impermeable to H^+, OH^-, K^+ and many other ionic solutes. The ATP formed inside the mitochondrial matrix will have to leave the matrix and reach the cytosol to become useful. Similarly any energy converting process requires a continuous supply of substrates and disposal of products and wastes. In several biological functions, products made at one part of the cell will have to be transported to another. Membrane transport, i.e. transport of molecules and ions across the membrane, is achieved in two general ways: (i) active transport and (ii) passive transport. For passive transport no energy input is required and the movement of the solutes is in the direction of the thermodynamic gradient. In active transport the movement is against the gradient and hence requires a source of energy. Membranes are made up of lipid bilayers, which divide the cell into compartments. They have a finite thickness ranging from 60 to 100 Å. They have selective permeability for certain ions and molecules. As the molecules will have to traverse through a hydrophobic lipid bilayer, there is good correlation between the lipid solubility of molecules and the penetration rate across membranes. Thus one would expect the rate of penetration of substances like ions, amino acids, etc. to be poor and this is indeed what is observed. However, this fact alone does not explain membrane transport completely, especially the high degree of selectivity observed.

A carrier is one of several mechanisms proposed for membrane transport (Figure 11.23). There are carrier molecules *A* in the membrane which have a high affinity for the species *B* to be transported. When *A* and *B* come together they form a complex *A* + *B*. Though *B* by itself has poor permeability, the complex *A* + *B* can traverse the membrane. For simplicity, we assume that the complex is electrically neutral. Valinomycin is an example of a carrier molecule. It is a cyclic peptide antibiotic with a hydrophobic exterior and a hydrophilic interior and is highly specific to K^+ ions. Other antibiotics from micro-organisms, like gramicidin and alamethicin are known as channel formers. As their name suggests, these molecules form channels across the membrane. The specie to be transported

binds at one end and travels through the channel to the other end. Channel former molecules do not move across the membrane to transport the ions. The interiors of the channel formers are hydrophilic and the exteriors are hydrophobic, just as in carrier molecules. Whether a particular mode of transport is active or passive is determined by the change in the free energy (ΔG) of transported species. If ΔG is positive it is active transport, and the mechanism requires a carrier molecule. If ΔG is negative it is passive transport and can occur spontaneously.

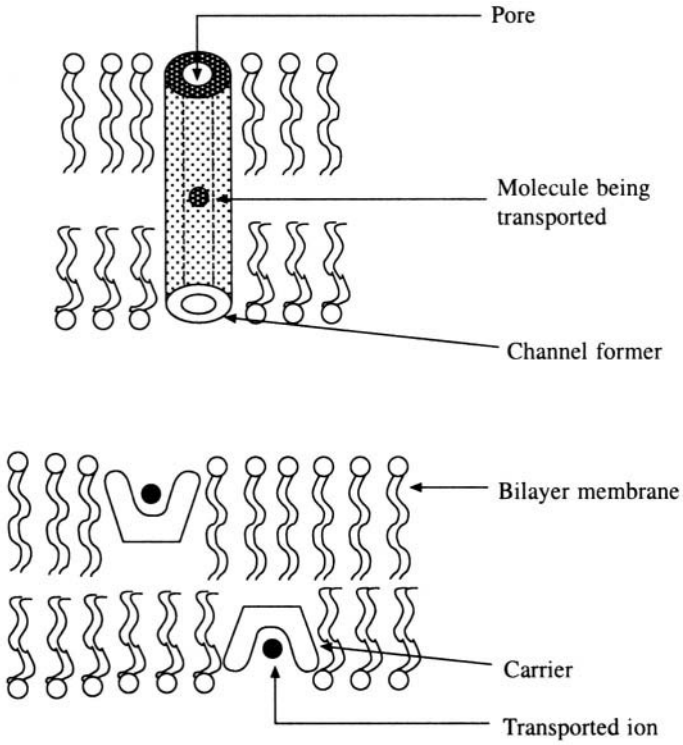


Fig. 11.23 Mechanisms of membrane transport

11.8.1 Active transport

In red blood cells, the cytoplasmic membrane is permeable to both Na^+ and K^+ ions but the K^+ ion concentration inside the membrane is greater than outside. For Na^+ the reverse is true. Thus an ionic gradient is maintained by a specific transport system known as the Na^+-K^+ pump. This ionic gradient is crucial for signal transmission through nerves. This difference in K^+ and Na^+ ion concentration can be maintained only by active transport and in fact the Na^+-K^+ pump maintains the gradient at the expense of hydrolysis of ATP. Thus active transport is coupled with the process that supplies energy. Under suitable conditions, this pump can be reversed to synthesise ATP from ADP and P_i. A model has been proposed for the Na^+-K^+ pump. The model takes into account the discovery, by Jens Skou, of an enzyme which hydrolyses ATP and which is present among the molecules that go to make up the pump mechanism. According to this model, hydrolysing enzyme in Na^+-K^+ pump must be in a position to adopt two conformations, and the affinity for transported species must be different for the two conformations. These are such that the cavity, or the site at which the species to be transported binds, is on the outside for one conformation and on inside for the other (Figure 11.24). From the figure we see that if E_1 has a higher affinity for Na^+ while E_2 has a higher affinity for K^+, the pump

would function in the required manner. As stated earlier, the action of ATPase is reversible and hence transport of ions across the membrane in the direction of thermodynamic gradient must cause the synthesis of ATP from ADP and P_i. In fact the proton gradient across the inner membrane of the mitochondria is linked to the production of ATP. The proof of this comes from the fact that no ATP is synthesised when the membrane is not intact.

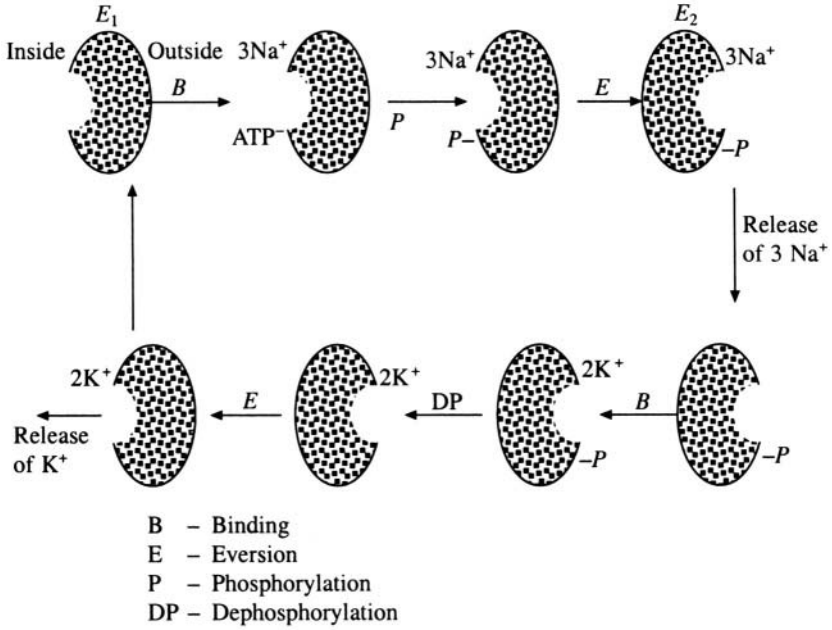


Fig. 11.24 Model of a Na^+-K^+ pump

11.8.2 Chemi-osmotic theory—Passive transport

As seen in a previous section, during electron transport in the inner membrane of the mitochondria, H^+ ions are ejected, generating a proton gradient across the membrane. According to the model proposed for this process (Figure 11.25), there are three H^+ conducting loops in the respiratory chain. Each loop translocates two H^+ ions from the inner matrix of the mitochondrion to the exterior through a carrier. The two electrons, which remain after the transport of H^+ ions are carried back to the matrix by a carrier. Thus for every pair of electrons passing from H_2 to oxygen, the loops carry $6H^+$ ions from the matrix to the external medium. Uncouplers like dinitrophenol (DNP) act by carrying protons across the membrane thus dissipating the proton gradient. This is necessary for ATP synthesis. In the case of chloroplasts, light causes water to loose electrons, which flow across the membrane and release protons in the inner medium. The reduced carrier on the other side of the membrane picks up protons from the outer medium and releases electrons to the second light reaction. The alternation of hydrogen and electron carriers on both sides of the membrane causes a net flow of protons across it. (Figure 11.25(b)). The thylakoid membrane and the mitochondrial membrane are permeable to water but not to hydrogen ions. Hence a concentration gradient of protons is formed and a proton motive force is built up. This proton gradient causes an electrochemical potential difference across the membrane which is given by

$$\Delta\mu = \mu_2 - \mu_1 = RT \log ([H^+]_2/[H^+]_1) + F(\psi_2 - \psi_1)$$

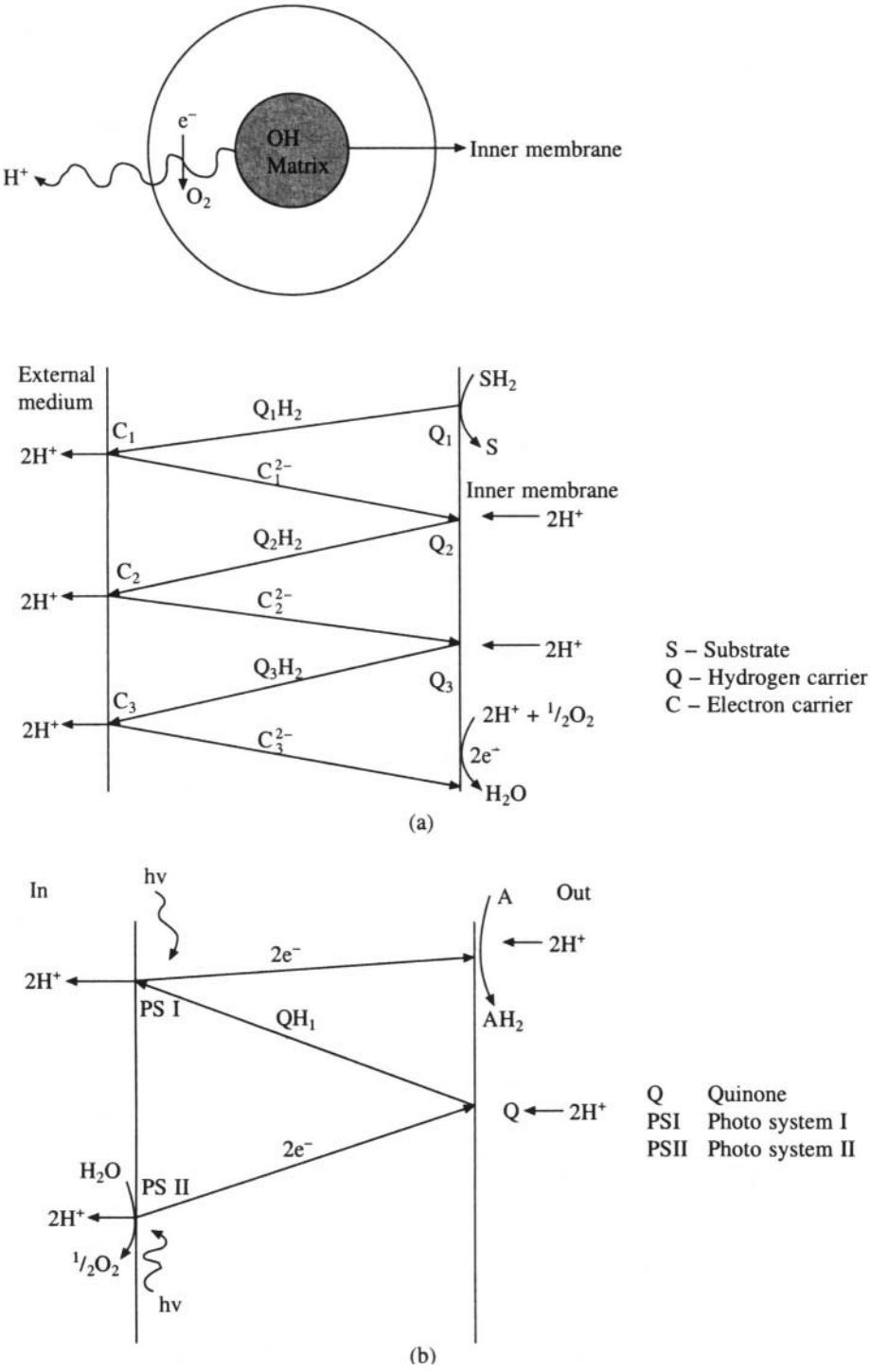


Fig. 11.25 Chemi-osmotic model for membrane transport in a: (a) mitochondrion and (b) chloroplast.

where F is the Faraday constant and $\Delta\psi = (\psi_2 - \psi_1)$ is the contribution from membrane electric potential. Since $pH = -\log [H^+]$, the above equation can be written as

$$\Delta\mu = 2.303RT(\Delta pH) + F\Delta\psi$$

where $\Delta pH = pH_1 - pH_2$ when hydrogen ions move from position 1 to position 2, and the factor 2.303 is due to conversion from natural logarithms to log base 10. Both ΔpH and $\Delta\psi$ are negative and that is the direction in which ions are pumped in. When the electrochemical potential is divided by the Faraday constant F we get the proton motive force or pmf. (analogous to the electron motive force, emf). In chloroplasts almost all the contribution to pmf is from the pH gradient. In mitochondria the contribution from the membrane potential is higher. This is due to the fact that the transfer of ions into the thylakoid space is accompanied by the transfer of either Cl^- in the same direction or Mg^{2+} in the opposite direction leading to neutrality. Hence no membrane potential is developed in chloroplasts.

12

Biomechanics

12.1 Introduction

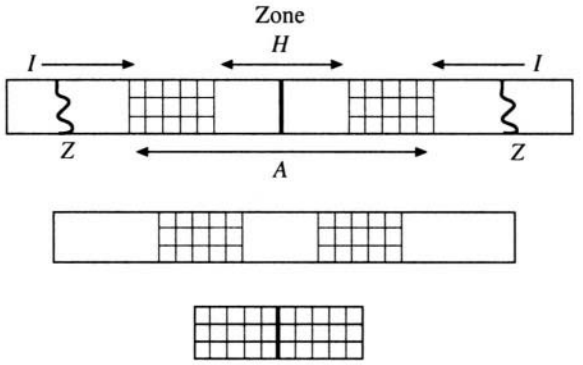
In this chapter we will deal with the conversion of chemical energy into mechanical energy. Mechanical work is done by all living organisms. It gives the organism the ability to move away from a noxious environment or move towards a beneficial region. There are movements inside the cell such as pinocytosis, cell division, etc. However, the mechanism involved in these situations is not well understood. This chapter focuses on larger systems such as the entire organism or the limbs, etc. The best known example in this system of conversion of chemical energy into mechanical work is that of muscle action. The proteins involved in these processes are actin and myosin and the chemical substance providing energy is adenosine triphosphate (ATP). Whenever ATP is converted to ADP (adenosine diphosphate) or AMP (adenosine monophosphate) approximately 8 Kcal/mole of energy is released, and is used for mechanical work by the muscle proteins.

Muscles can be classified as smooth muscle, cardiac muscle and striated or skeletal muscle. The walls of some internal organs, like the bladder, the esophagus, intestines, etc. are made of smooth muscles. The contraction of smooth muscles is not under voluntary control but takes place in a slow, generalised manner. The cardiac muscles are fibrous and also slightly striated. They undergo periodic contraction and relaxation automatically. Striated muscles are under voluntary control. They are highly organised and consist of several elongated cells. These elongated cells or muscle fibres in turn contain the myofibrils. In a light microscope the muscle cells show a characteristic repeating structure which gives it a striated appearance.

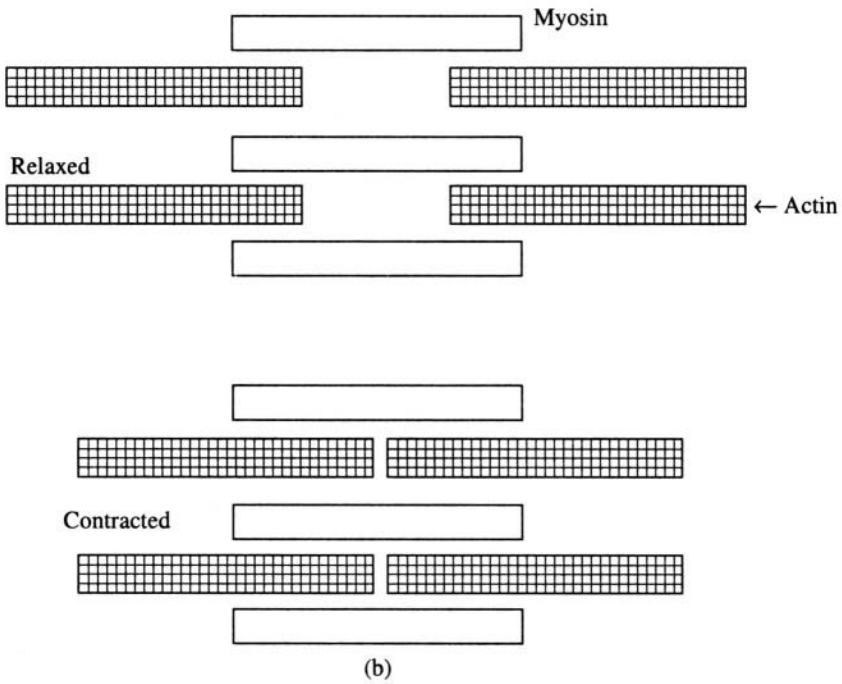
12.2 Striated Muscles

Striated muscle cells are multi-nucleated and are enclosed in an electrically excitable membrane called the sarcolemma. The intracellular fluid surrounding the myofibrils is sarcoplasm. It contains glycogen, ATP, phosphocreatine and glycocyte enzymes. When a longitudinal section of myofibrils

is examined under an electron microscope one can see the repeating unit, sarcomere, every **2.3 μm** along the fibre axis. A sarcomere has a banded appearance due to regions with different optical properties. A dark *A* band and a light *I* band alternate regularly (Figure 12.1). The middle of the A band is known as the *H* zone and a dark line is found in the middle of the *H* zone. The Z lines bisect *I* bands. There are two kinds of filaments in the sarcomere. The thick filaments make up the *H* zone of the *A* band and only thin filaments are found in the *I* band. These filaments are protein filaments of diameters 150 Å and 70Å respectively. The thick filament consists mainly of the protein myosin while the thin filament consists of actin, tropomyosin and troponin. Each thick filament is surrounded by six thin filaments in the region of overlap, and each thin filament has three neighbouring thick filaments. The filaments interact with each other through what are known as cross bridges. Electron



(a)



(b)

microscopy and optical measurements show that the *I* zone shrinks when the muscle contracts. Based on this, a 'sliding filament' model of muscle action has been proposed. According to this model, the lengths of the filaments do not change but the length of the sarcomere decreases as the filaments slide past each other during contraction. Thus, though the lengths of the thick and thin filaments do not change, the sizes of the *H* zone and the *I* band decrease. The cross bridges change angle during the sliding motion of the filament. When at rest the thin and thick filaments overlap only to an extent of about one third of their length. At maximal contraction, the thick and thin filaments fully overlap and shortening is about 32 Å (Figure 12.1(b)).

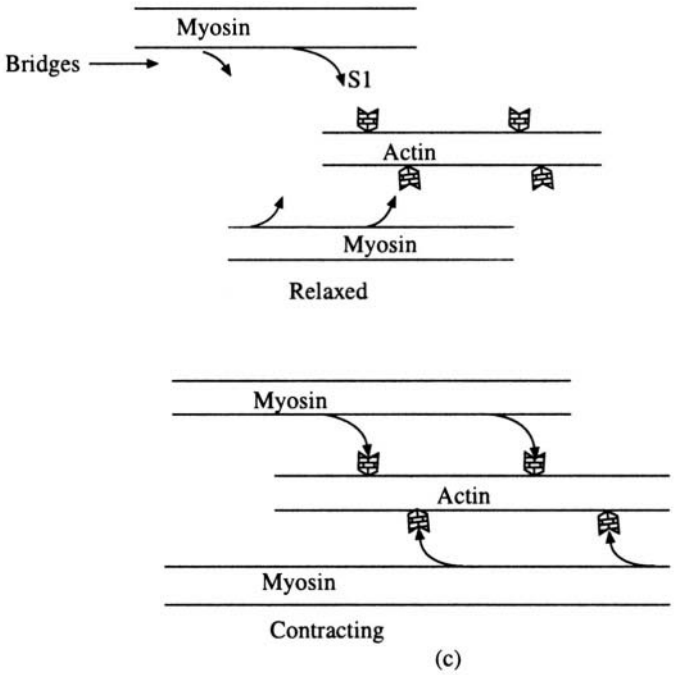


Fig. 12.1 A schematic representation of the sliding model of muscle action. (a) As seen under a microscope. (b) Thick and thin filaments in a contracting muscle. (c) Formation of coss bridges

12.2.1 Contractile proteins

The major components of myofibrils are the protein myosin (50 to 55% of the total protein) and actin (20 to 25%). Actin is the major constituent of thin filaments. When extracted from muscle, a globular protein known as *G*-actin emerges. In the presence of Mg^{2+} and ATP *G*-actin polymerises to form the fibrous *F*-actin. *F*-actin looks like two strings of beads wound round each other when looked under electron microscope. The other two proteins that are closely associated with actin in thin filaments are tropomyosin and troponin. Tropomyosin is a double stranded **α-helical** rod arranged parallel to the long axis of the thin filament. Each tropomyosin molecule covers about seven *G*-actin monomers. Troponin is a complex protein with three subunits viz. TnC, TnI and TnT having the molecular weight of 18, 24 and 37 kDa respectively. Troponin is attached to the tropomyosin filament, one on each molecule. TnC binds calcium ions, TnI binds to actin and TnT binds to tropomyosin. The troponin-tropomyosin complex plays an important role in the regulation of contraction, which is effected by Ca^{2+} ions (Figure 12.2).

The thick filament consists mainly of myosin molecule whose molecular weight is about 500 KDa. Electron microscopy shows that myosin has a double headed globular protein attached to a long rod. The rod is a two-stranded **α-helical** coiled coil. When myosin molecule is cleaved using the enzyme trypsin, it splits into two fragments, known as light meromyosin (LMM) and heavy meromyosin. The heavy meromyosin molecule, which contains the globular heads, has ATPase activity while the light meromyosin lacks this activity. The heavy meromyosin (HMM) does not form filaments (Figure 12.3). The globular heads also have a strong affinity for actin suggesting that this part of the myosin molecule must be involved in forming the cross bridges with actin in the myofibril. The formation of this complex leads to a large increase in viscosity, but this is reversed on addition of ATP. Figure 12.3(a) is a schematic diagram of the myosin molecule showing the six-polypeptide chains that form this macromolecular assembly. Myosin has a molecular weight of about 500 KDa and is made up of six polypeptide chains, viz. two heavy chains (= 400 KDa) and two pairs of light chains (each with a molecular weight of about 20 KDa). Each one of the myosin heads is globular in nature and has about 850 amino acids contributed by the respective heavy chain and two light chains. The remaining portions of the two heavy chains assemble in an extended coiled coil form, which is about 1500 Å long.

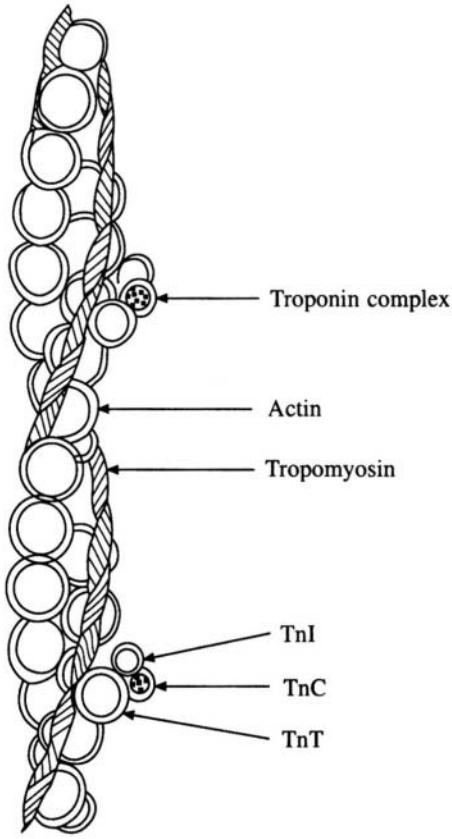


Fig. 12.2 Model for a thin filament in resting stage

12.3 Mechanical Properties of Muscles

Muscles are transducers capable of converting chemical energy into mechanical energy. In a typical muscle function a load is moved through a distance due to shortening of the muscle fibres. The contractile unit of a muscle fibre is known as a tension generator. A muscle is elastic and develops tension when it is stretched. When a muscle is greatly stretched, the tension increases to a point beyond which muscle fibres can break. The elastic system of a skeletal muscle basically consists of three components, a contractile component and two elastic components. Sarcoplasm and sliding filament constitute the contractile component and tendons, disc, connective tissues and sarcolemma form the elastic components. Experimental observations of muscle contraction can be done using thermal, mechanical or electrical stimulus. However, electrical stimulus is preferred as the muscle is not damaged in the process and the intensity of the stimulus can be accurately measured using a galvanometer. A muscle twitch, which is a simple response to the threshold stimulus, is useful in the study of properties of muscles. A muscle switch goes through three stages: (1) latent period, (2) contraction phase and (3) relaxation phase. A muscle fibre can continue to be in tension if a series

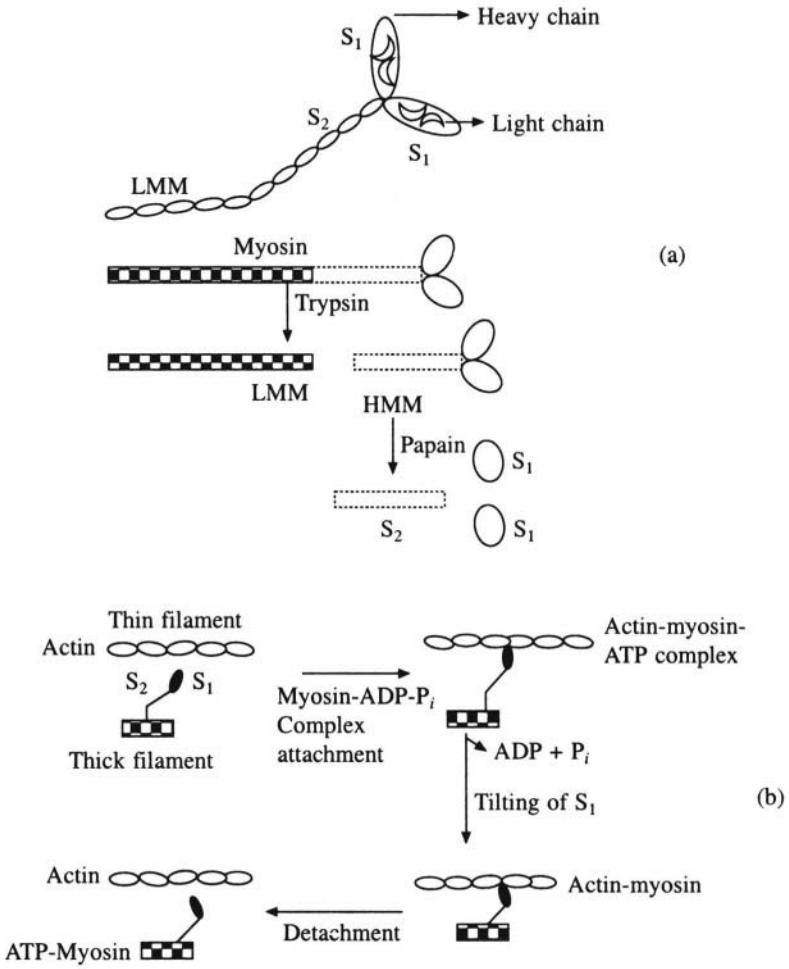


Fig. 12.3 (a) Schematic diagram of myosin molecule. (b) Proposed mechanism of muscle contraction

of stimuli are applied at high frequency without allowing the muscle to relax. This is known as tetanus. A fibre capable of maintaining tetanus for a considerable time can stop contraction due to. fatigue if the tetanus is prolonged too long.

12.3.1 Contraction mechanism

The decrease in viscosity when ATP is added to the troponin-tropomyosin (or actomyosin) complex undergoes slow recovery when ATP is hydrolysed. This is because ATP is required for breaking the cross bridges between actin and myosin rather than for making it. Thus the steps involved in muscle contraction are cyclic and can be designated as the following.

1. Binding of ATP to the myosin head (also known as $\mathbf{S_1}$ head).
2. Activation of myosin ATP complex.
3. Binding of the activated myosin-ATP complex to actin in thin filaments in the presence of $\mathbf{Ca^{2+}}$ ions.

4. ATP hydrolysis followed by the change of orientation of S_1 heads with respect to the filament.
5. Binding of ATP to myosin heads leading to the dissociation of myosin-ATP complex from actin filament.
6. The cycle repeats from step 2.

Thus the above model proposes two forms for the actomyosin complex: (1) The activated form. (2) The inactivated form. The binding of myosin-ATP complex to thin filaments in the presence of Ca^{2+} ions is the activated form. The part that is left after hydrolysis of ATP is the inactivated form. Figure 12.3(b) shows the details of the proposed mechanism of muscle contraction. The contractile force is generated by the making and breaking of the complexes between *S* heads of myosin and actin. In the resting stage, S_1 heads are detached from thin filaments, and myosin contains tightly bound ADP and P_i. When muscle is stimulated, S_1 heads move away from thick filaments, and get attached to the thin filaments in a perpendicular orientation. The tightly bound ADP and P_i are released and the orientation of S_1 head changes to make an angle of 45° with the thin filament axis. This tilt of the S_1 head is the power stroke of muscle contraction. The tilt is then transmitted to the thick filament through the S_2 unit (rod region) of the myosin molecule. This results in pulling the thick filament by about 75 Å with respect to the thin filament. In the final step, the S_1 head is detached from the thin filament by the binding of ATP and reverts back to its perpendicular orientation. Finally, hydrolysis of ATP by S_1 completes the cycle. Two types of hinges are invoked in the contractile mechanism—one between the S_1 and S_2 units, and the other between S_2 and light meromyosin.

12.3.2 Role of Ca^{2+} ions

It has been shown that the onset of contraction is mediated by Ca^{2+} ions. The contraction of muscle is initiated by the arrival of a nerve impulse at the end of the muscle fibre and the consequent release of acetylcholine. The muscle membrane (the sarcolemma) has a resting potential of 70mV. The response of the membrane is a change in membrane permeability, with Na^+ ions flowing into the muscle and K^+ ions out of it, and the consequent change in membrane potential. Within a few milliseconds the cell begins to contract as the local response propagates over the membrane. The nerve excitation triggers the release of Ca^{2+} ions by the sarcoplasmic reticulum. The released Ca^{2+} ions bind to TnC component of troponin and causes conformational changes, which are transmitted to tropomyosin and actin. Consequently tropomyosin moves towards the helical groove of the thin filament which enables S_1 heads of myosin molecules to interact with actin unit of the thin filament. Contraction takes place with concomitant hydrolysis of ATP till Ca^{2+} ions are removed and this causes the S_1 head to be blocked by tropomyosin again. Thus the binding of Ca^{2+} to troponin causes the activated myosin-ATP binding site on actin to be exposed. This is otherwise blocked by troponin and tropomyosin (Figure 12.2). The flow of information can be summarised as

$$Ca^{2+} \rightarrow \text{Troponin} \rightarrow \text{Tropomyosin} \rightarrow \text{Actin} \rightarrow \text{Myosin}$$

Though it is known that ATP is the source of energy for muscle contraction, the supply of ATP is through a phosphocreatinine which has higher phosphate group transfer potential than ATP.

$$\text{Phosphocreatinine} + \text{ADP} \Leftrightarrow \text{ATP} + \text{Creatine}$$

However phosphocreatine itself cannot make muscles contract.

The experiments on muscle contraction are performed by applying the potential directly to the tissue. Single fibres respond only when the potential exceeds a threshold and hence are not preferred. Two types of experiments on muscles can be carried out. In one case, both the ends of the muscle are fixed so that contraction occurs without shortening. This condition is known as isometric. In the

second case only one end of the fibre is fixed and hence the muscle contracts and shortens and this condition is known as isotonic. A.V. Hill carried out fundamental studies on muscle behaviour and he showed that the maximum tension exerted by a muscle is about **5 Kg-m/cm^2** of the muscle cross section. Figure 12.4 shows how the mechanical advantage and efficiency of muscle contraction may be calculated.

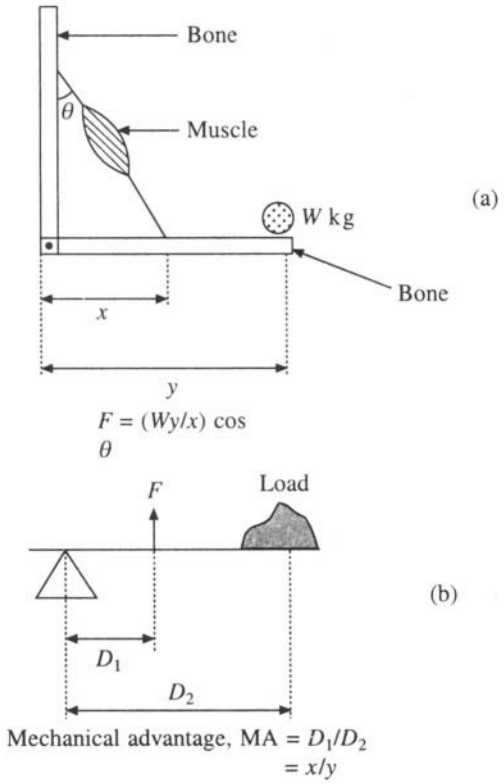


Fig. 12.4 (a) Force exerted by a muscle. (b) Mechanical advantage of the arm

12.4 Biomechanics of the Cardiovascular System

The blood circulatory system is a transport system that transports gases like O_2 and CO_2, nutrients, wastes, hormones and immunologically active substances. Blood flow can be considered as a more or less closed circulatory system, which is pumped by the heart. (It is not a completely closed system since some fluid, called lymph, leaves the capillaries and flows into the tissue spaces). Blood is a suspension of different types of single cells in a viscous medium containing proteins, organic salts etc. Kidneys, heart and lungs are the major organs that interact with blood. The vessels into which blood is pumped by heart are known as arteries. They branch into smaller and smaller arteries and finally form capillaries. Most of the exchanges between blood and the surrounding tissues take place at the capillary level. The capillaries then join together to form venules, leading to larger and larger veins which bring back the blood to the heart. The kidney is the most crucial organ. It removes waste products and excess water from the blood, thus maintaining the blood volume, ionic concentrations, arterial pressure etc. Thus the circulatory system is in a dynamic state, losing water to kidneys, lungs, exocrine organs (e.g. salivary glands), etc., and re-absorbing it from the gut and also that formed during metabolic processes.

12.4.1 Blood pressure

The fundamental properties of any fluid are pressure, velocity and density. Pressure is force per unit area, and is generally measured in terms of the height of a column of liquid that can be supported. The most frequent system of units used to measure blood pressure is millimetres of mercury, being the height of the column of the liquid metal supported by blood pressure. The maximum arterial pressure is called the systolic pressure and the minimum arterial pressure is called the diastolic pressure. The pressure falls when the blood reaches the capillaries. At the venous system it is still lower. The arteries and veins have very similar inner diameter (0.5 to 12.5 mm) and flow rates, but the pressures are different. This is due to the fact that arteries have thick, elastic walls while veins have thin walls. The larger pressures in arteries prevent any reverse flow of blood from veins. The pulsations in the arteries are due to the stretching of its elastic walls from the force of the heartbeat.

The mammalian heart is shown in Figure 12.5. It consists of four chambers, two auricles and two ventricles. Blood from the entire body, excepting from the lungs, enter the heart at the right auricle. From there it is pumped into the right ventricle, then into the lungs and finally back into the left auricle. From the left auricle the blood goes to the left ventricle and from there through the aorta (which is the largest blood vessel leading away from the heart) to the arteries and is then distributed to the entire body, excepting the lungs. This system of circulating blood is very efficient in supplying oxygen and removing carbon dioxide. Four valves viz. the tricuspid valve and mitral valve separating the auricles and ventricles, and the pulmonary valve and semilunar valve prevent the back flow of the blood from the ventricles to the arteries or from the main arteries into the respective ventricles (Figure 12.6).

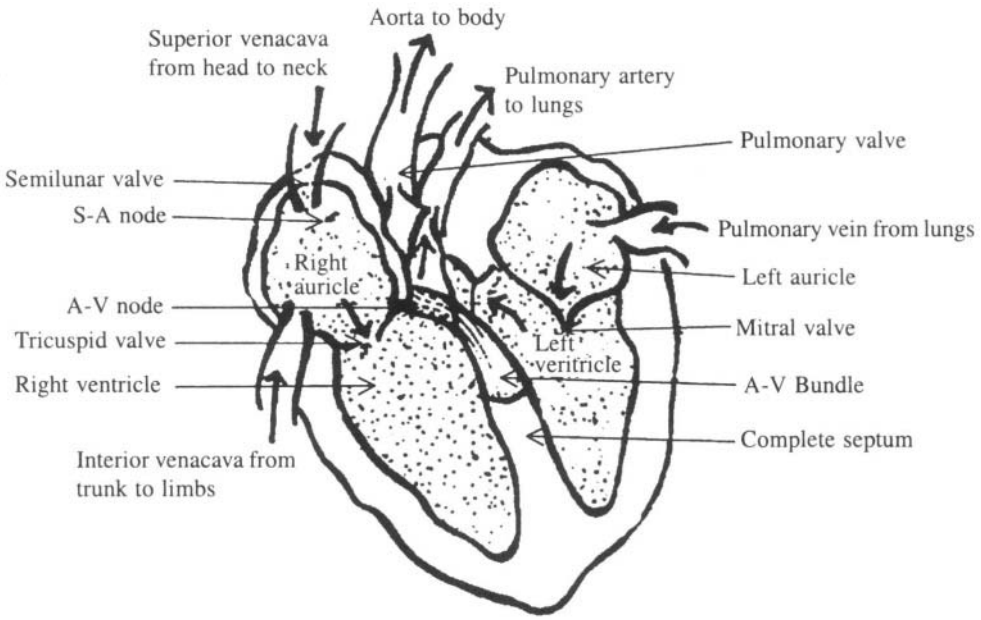


Fig. 12.5 Mammalian heart. Thick arrows show direction of blood flow.

The mechanical action of the heart is due to the heart muscle (myocardium) and is analogous to those of the skeletal muscle. As in other striated muscles the cardiac muscle fibre membranes are also polarised, i.e. they have 90 mV negative potential between the inside of the membrane and the outside. Just as in nerve fibres, the polarity reverses for a short period of time, when an impulse is received by the membrane, just before contraction occurs. The action potential is around 120 mV, i.e. the outside will be 30 mV more negative than the inside at the peak of the spike. The large collection of cardiac muscle fibres acts like a group of electric cells and leads to measurable potential changes on the body surface. These potential differences can be measured using an electrocardiogram. The heartbeat is initiated by a sino-auricular (*s-a*) node and the node acts like an electronic multivibrator letting out one electrical pulse per heart cycle. The heart pulses are periodic and rhythmic. During systole, which is the active part of the heart cycle and when the arterial pressure is at a maximum, the blood from ventricles is forced into aorta and pulmonary artery respectively. In the beginning of this cycle all the four valves, are closed. As the pressure increases, and the ventricular pressure just exceeds the pressure at the roots of the aorta or the pulmonary artery, the semilunar valve opens. The

ventricular pressure reaches a maximum and gradually decreases and this causes the semilunar valve to close. The second part of the heart cycle known as diastole then begins with the relaxation of the ventricular wall and a drop in the ventricular pressure with respect to the corresponding attrium. Hence the atrio-ventricular valves (*a-v*) open and the blood enters the ventricles. At the end of the diastolic cycle the atria contract and the pressure inside the ventricles increases. Immediately after this the ventricular valves close and the systol and diastole cycle is repeated.

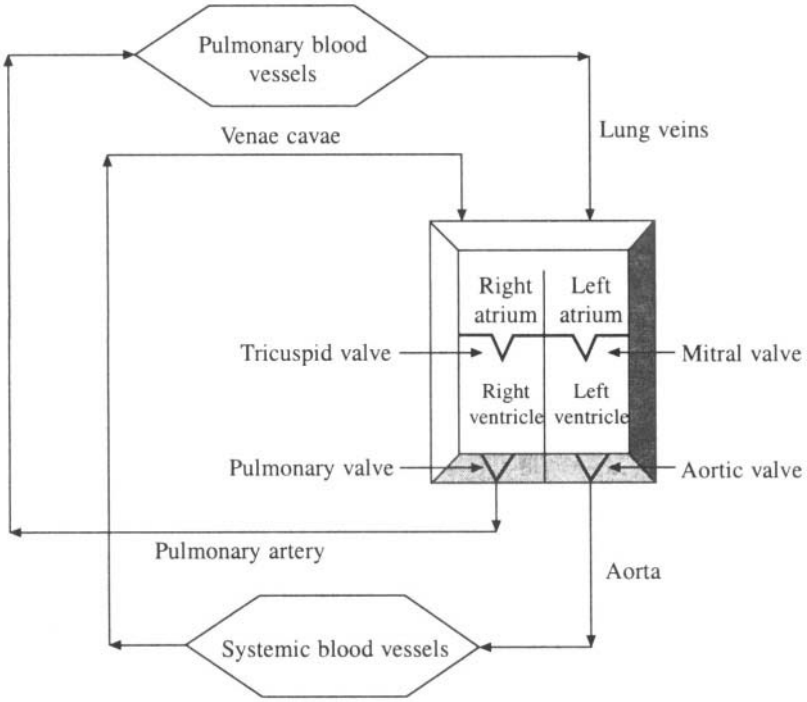


Fig. 12.6 Circulation of blood

12.4.2 Electrical activity during the heartbeat

The heart pulses generated by the *s-a* node spread over the surface of the auricle and make the muscle fibres to contract. In addition, the electrochemical pulses stimulate the *a-v* node. After a time delay of 0.1 seconds or less, this node lets out a new electrical pulse that is carried by special fibres known as *a-v* bundle to the septum. These fibres terminate at different points in the septum and on the outer ventricular wall. Thus the fast conduction of the impulses by *a-v* bundle synchronises with the ventricular contraction. The (*a-v*) node can take over the control of heart rate when the (*s-v*) node has failed. Under these circumstances auricular contraction will not be properly synchronised with ventricular action. If the *a-v* node also fails, the auricular and ventricular walls take over the control of heart rate. These conditions are not fatal.

The difference between cardiac muscle fibres and nerve fibres lies in the recovery process. The recovery of resting potential in nerve axons takes about a fraction of millisecond to 5 milliseconds. On the contrary, the cardiac muscle fibres take as long as 200 milliseconds to recover their resting potential. Similarly, there is a subtle difference in the permeability of $\mathbf{K^+}$ ions during and after the action potential in the case of cardiac muscles as compared to nerve axons. This was seen by

experiments using squid axons. The permeability of K^+ ions increases suddenly and then drops when the resting potential of a voltage-clamped squid axon is suddenly decreased. In the case of the cardiac muscles, the original low or zero permeability to potassium ions is recovered only after the membrane potential returns to its original value.

12.4.3 Electrocardiography

As mentioned above, electrical potential changes occur during heartbeats and this is spread over the surface of the body. Electrodes at any two points of the surface of the body will be able to measure this potential difference and it can be recorded using an electrocardiogram (ECG). As the heart is not the only organelle capable of generating such potentials at the body surface, adequate precautions will have to be taken while recording an ECG. Even a movement of a skeletal muscle can give rise to body surface potential difference and hence the patient is made to lie down during recording of the ECG. The electrodes are placed generally at the right arm and left leg. Sometimes three wires are attached, one to each arm and the third to left leg. The ECG is then recorded between each pair of electrodes. The maximum potential difference is about 1 mV and this is adjusted to show 1 cm deflection on the ECG. The electrical signals corresponding to one heart cycle, also known as the waves of the signal, are shown in Figure 12.7. The *P* wave occurs just before the auricular contraction and the QRS complex occurs at the start of the ventricular contraction. The end of ventricular contraction gives rise to *T* waves. The ECG cycle is described in Table 12.1

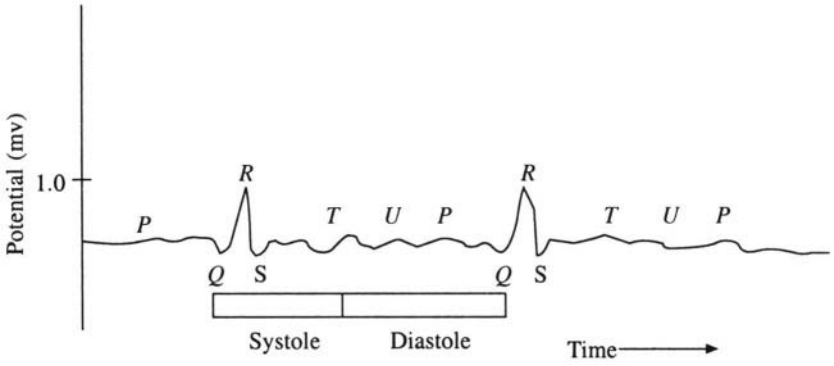


Fig. 12.7 Electrocardiogram. P wave precedes auricular contraction, QRS is due to ventricular contraction

Table 12.1. The ECG cycle of signals and the corresponding cardiac functions

ECG waves	Duration (milliseconds)	Corresponding heart cycle
P	8	Precedes auricular contraction
P–Q	150–200	*a-v* delay time
Q	40–80	Precedes ventricular contraction
R		
S		
S–T	100–250	Ventricular ejection
T	100	Follows ventricular ejection
T–P	300	Diastole

The potential difference between two points can be written as

$$V_1 = V_{LA} - V_{RA}$$

$$V_2 = V_F - V_{RA}$$

$$V_3 = V_F - V_{LA}$$

(The subscripts LA, RA and F refer to the left arm, the right arm and foot respectively. Instead of the foot, the chest or the back can also be used for potential measurements.) Therefore

$$V_2 = V_3 + V_1$$

This is known as Einthoven's Law. Thus if any two of the three potential differences (also known as leads in ECG terminology) are measured the third can be calculated. Further if V_1, V_2 and V_3 are taken to be vectors, they are chosen such that

$$V_{LA} + V_{RA} + V_F = 0$$

taking into account the signs of the potentials (Figure 12.8). The triangle thus constructed is known as Einthoven's triangle. Einthoven further showed that the three vectors could be considered as projections of a single hypothetical heart vector **H** on this triangle. Though this relation is true for any three points, the **H** vector will be meaningful only when it is related to the electrical axis of the heart. The three points chosen are such that they are equidistant from the heart so that the triangle is an equilateral one.

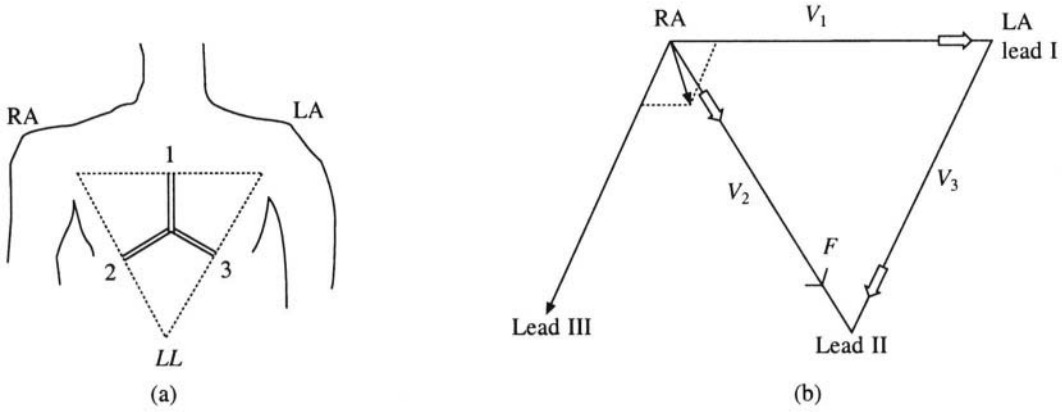


Fig. 12.8 Einthoven's triangle

A realistic model for the electrical events of the heart cycle is based on the assumption that the heart acts as a single dipole, represented by the **H** vector, and the torso is a homogeneous conductor. The recording of the magnitude and spatial orientation of the equivalent heart dipole as a function of time is known as vector cardiography. However, in practice it is not possible to locate a single equivalent dipole and to monitor its magnitude and orientation with respect to time. One of the ways of improving the performance of Vector ECG is to increase the number of electrodes used in order to get a refined value of the dipole. Clinical ECG tests may therefore involve use of 12 or more leads to aid in diagnosis In spite of its great utility, many of the abnormalities in the functioning of the heart may not be obvious in vector cardiography. A more appropriate method would be to represent heart as a collection of dipoles. By the use of such techniques and of computer based automated diagnosis, vector cardiograms have proved clinically very useful.

13

Neurobiophysics

13.1 Introduction

The external world is perceived by organisms through stimuli received by them. These can be mechanical, thermal or chemical. They are transduced by the sensory receptors of the organism into electrical signals and passed on to the central nervous system. The internal communications within the organism are taken care of by the nervous system. This transmission of information relating to the environment and the ability of multi-cellular organisms to respond appropriately to them is essential for survival. This chapter deals with the nerve transmission and other forms of information transport.

As far as the human beings are concerned vision, hearing and smell play vital roles in communications. Hence, the mechanism of hearing and vision is discussed in some detail here. Chemical signals, like the odours emanating from flowers, are described in the later part of the chapter.

13.2 The Nervous System

The nervous system consists of neurons, which transmit information from one part of the organism to the other. This is known as the neuronal response. The nerve cells (neurons and Swann cells) have a cell body from which small projections known as dendrites and large projections called axons protrude. The dendrites carry impulses towards the central cell body (Figure 13.1). The cell body is the thick region of the neuron containing the nucleus and most of the cytoplasm. The axons carry impulses away from the cell body. A fatty covering known as the myelin sheath covers most of the larger axons. The myelin sheath is interrupted somewhat periodically at sites called the nodes of Ranvier. Several axons are grouped together in each nerve. Similarly nerve cell bodies form clusters known as ganglia. Some axons are very long and can extent up to 3 or 4 metres. For example the nerve controlling the muscles in the human finger extend up to the spinal cord where they have their cell bodies. The diameter of the axons is of the order of 1 to 100 μm. Connections between neurons are called synapses and these occur at the branched ends of axons, dendrites and collateral branches of

different neurons. Axons can also end in a synapse at the receptors such as the hair cells of the organ of the corti in the ear.

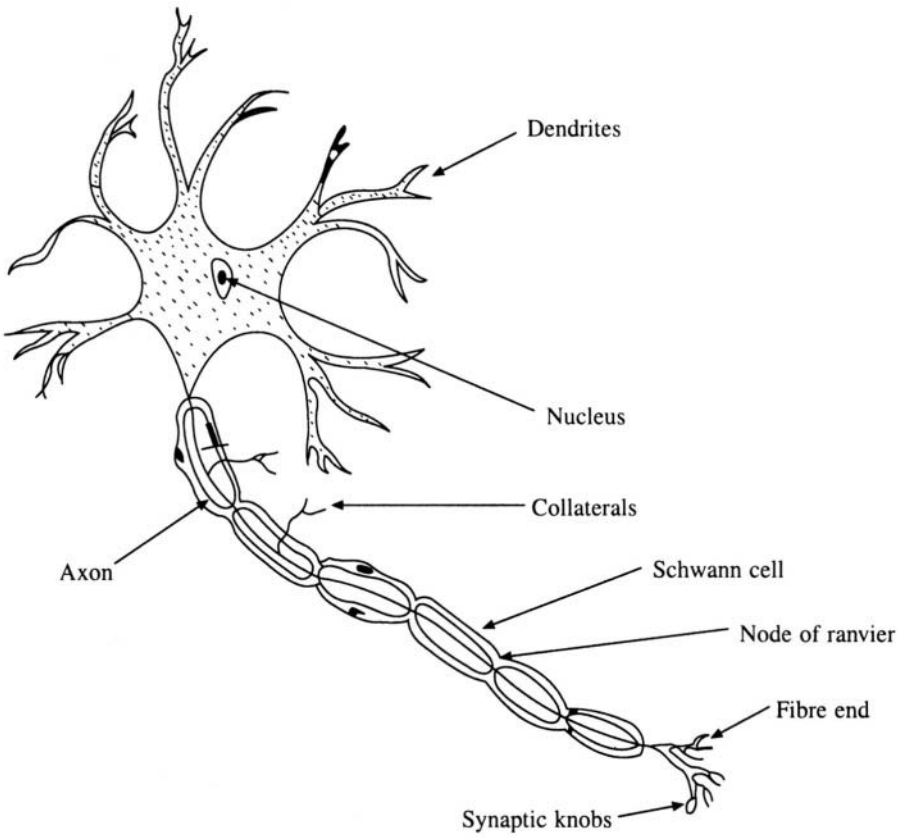


Fig. 13.1 Nerve cell

The nervous system of vertebrates is organised into the central nervous system and the peripheral nervous system. The central nervous system is in the skull and the spinal cord while the peripheral nervous system consists of nerves and ganglia. Generally the peripheral nervous system contains both sensory axons and motor axons. The sensory axons conduct signals toward the central nervous system while motor axons conduct them away from it. Though the neurons within the central nervous system look similar to those outside it, there are several differences. For example when a nerve fibre within the central nervous system is injred, the entire neuron degenerates. On the contrary if a peripheral nerve is cut, the fibres will re-grow out of the old nerve trunk.

A single axon can be removed from an organism and be examined in the laboratory. The maximum diameter of a vertebrate axon is about 20 μm while some of the invertebrate axons can be as large as 200 μm in diameter. Squid axons are easy to work with, due to their size, and are used routinely for laboratory studies. It is now possible to insert a microelectrode inside a nerve axon and to measure the electric potential relative to the surrounding medium (i.e. the interstitial fluid outside the fibre). This potential normally ranges from -50 to 100 mV. Axons can be stimulated by electrical pulses, heat, chemical changes and also mechanical pressures. When the axon is stimulated its surface potential changes in a specific way, producing a signal called the action or spike potential. The spike

potential thus formed is an all-or-none response and is transmitted down the axon in both the directions. However, due to the nature of the synapse present in an intact nerve, only one of these directions is usually effective. Each spike generally starts with a sudden change from its normal negative resting potential to a positive membrane potential, and then abruptly falls back to the resting potential. The total duration is about one millisecond (Figure 13.2).

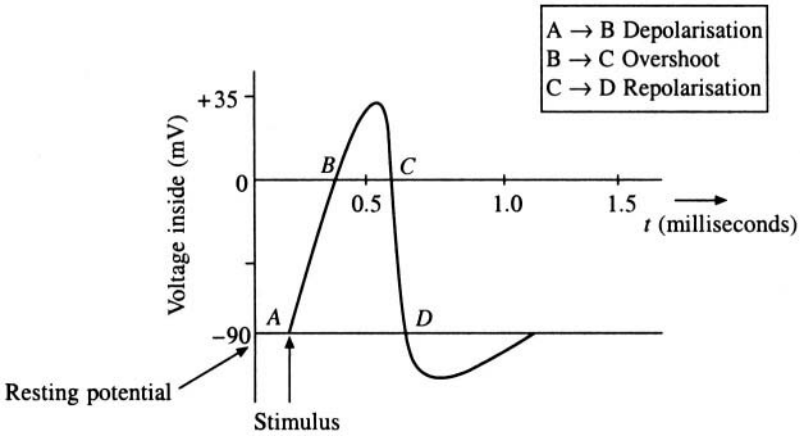


Fig. 13.2 A typical action potential

13.2.1 Synapse

In the nervous system the message is transmitted from one point to another. This means that a two-way passage of the message along the axons has to be prevented by some means. This is achieved by the synapse. The point at which an axon and a dendrite associate is called a synapse (Figure 13.3). When an action potential arrives at the presynaptic fibre, transmitter molecules are released from the vesicles and these diffuse through the synaptic gap and reach the sensitive surface of the postsynaptic fibre. The transmitter molecules alter the permeability to Na^+ and other monovalent ions in the membrane of the sensitive surface, thus creating an action potential. The transmitter molecules are

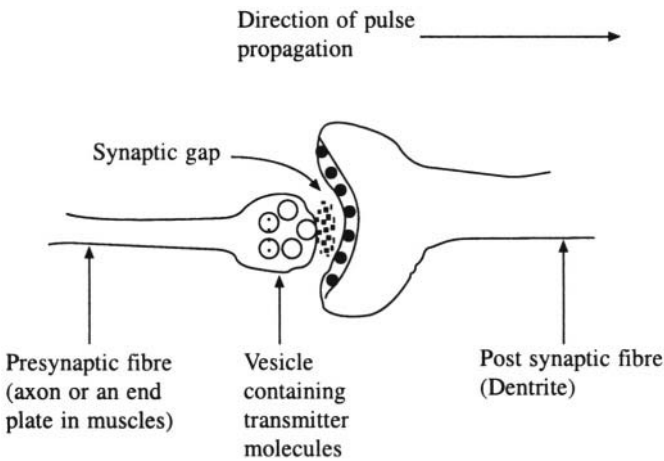

Direction of pulse propagation
Synaptic gap
Presynaptic fibre (axon or an end plate in muscles)
Vesicle containing transmitter molecules
Post synaptic fibre (Dentrite)

Fig. 13.3 A synapse

destroyed immediately after this so that the synapse could be ready to receive the next signal. Acetylcholine and noradrenaline have been identified as synaptic transmitters and a very small quantity $(10^{-18}\ M)$ of the substance is enough to evoke an action potential. This amount is known as threshold quantity. The destruction of acetylcholine is achieved by the enzyme acetylcholine esterase, which is situated in the receptor surfaces of the postsynaptic fibres. Synaptic conduction can also occur through charge transfer without any chemical intermediates. In this case the impulses travel from one axon to another electrically with negligible time delay. One spike potential at the presynaptic fibre can induce one spike potential at the postsynaptic fibre. In some cases, when the response due to spike is below the threshold value, then sub-threshold responses from several synapses can be added together to form a spike potential at the post synaptic fibre. Thus the neurons can act like a digital computer and can add, subtract or do other arithmetical operations.

13.3 Physics of Membrane Potentials

There are two ways by which membrane potentials can develop: (1) Diffusion of ions through the membrane causing an imbalance of negative and positive charges on the two sides of the membranes (Figure 13.4). (2) Active transport across the membrane again leading to imbalance of the charges on either side of the membrane. Measuring the membrane potential of a nerve fibre is a difficult task due to the small size of the fibres. A microelectrode with a diameter of 1 micron is inserted through the membrane. The other electrode, called the indifferent electrode, is kept in the interstitial fluid (Fig 13.5). The potential difference is measured using a sophisticated voltmeter, which is capable of measuring even very small voltages despite extremely high resistance to electrical flow through the tip of the microelectrode.

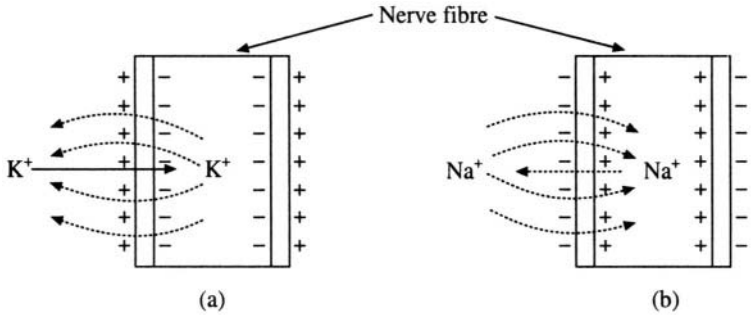


Fig. 13.4 Diffusion potential across cell membranes. Nernst potential for (a) K^+ ions and (b) Na^+ ions.

13.3.1 Membrane potential due to diffusion

Generally the K^+ ion concentration is greater on the inner side than on the outer side of the membrane. The reverse is true for Na^+ ion concentration. When there is no active transport, and if we assume that the membrane is permeable only to K^+ ions, then these ions will diffuse from inside to outside the membrane (Figure 13.4(a)), since the K^+ ion concentration inside is greater. The K^+ ions carry the positive charge to the outside and hence a state of electropositivity outside and electronegativity inside is developed. Thus a potential difference across the membrane is developed and this prevents further K^+ ions moving outward. The potential across the membrane is now known as the Nernst potential for potassium ions. Figure 13.4(b) shows the same effect for Na^+ ions, where we now assume that the membrane is highly permeable to Na^+ ions and impermeable to all other ions. As the

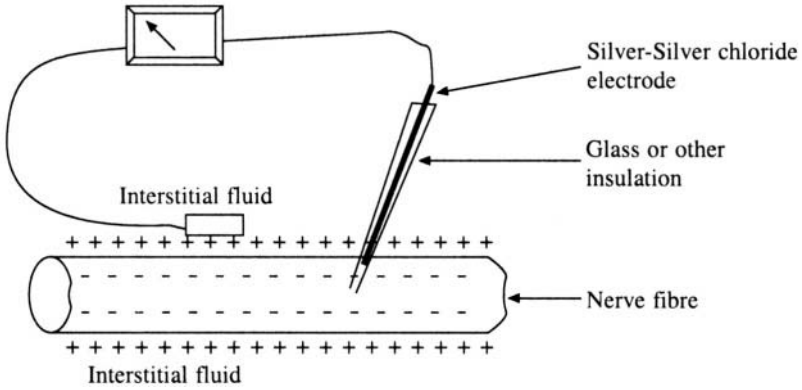


Fig. 13.5. Measuring membrane potential

Na^+ ion concentration outside is more, sodium ions move inward and a membrane potential with reverse polarity is generated. When the membrane potential rises high enough to prevent further diffusion of Na^+ ions, it is known as the Nernst potential for sodium ions. The magnitude of the potential is given by

$$\text{EMF(in mV)} = -61 \times \log \left[\frac{\text{Concentration of positive ions (inside)/}}{\text{Concentration of negative ions (outside)}} \right]$$

When the concentration of positive ions inside is ten times that of negative ions outside, then log (10) is 1, and the EMF = -61 mV. The above equation is called Nernst equation. For sodium ions, the Nerst potential is $\cong +61$ millivolts. For potassium ions, the Nernst potential $\cong -94$ millivolts. Under resting conditions the membrane potential averages to –90 millivolts which is near the Nernst potential for potassium. This is because at resting stage the membrane is more permeable to K^+ and only slightly permeable to Na^+ ions.

We can now see more details of how the action potential, mentioned above, develops across the membrane. There are three stages.

(a) *Resting potential:* This is the stage before the action potential actually occurs. The membrane potential is negative and is said to be polarised during this stage.

(b) *Depolarisation stage:* At this stage, the membrane becomes very permeable to sodium ions and a tremendous number of sodium ions rush into the interior of the axon. The membrane potential rises in the positive direction and sometimes overshoots the zero value.

(c) *Re-polarisation stage:* Sodium channels close immediately after this and potassium ions flow to the exterior re-establishing the resting stage negative potential (Figure 13.2). Voltage gated channels and Na^+–K^+ pumps play a vital role in both depolarisation and re-polarisation of the nerve membrane during the action potential. These are membrane proteins or peptides or molecular complexes, which open and shut appropriately to allow or prevent the flow of the ions. As the names suggest, the gates play a passive role, while the pumps are proactive. The sodium channel is activated when the membrane potential rises from its resting value of –90 mV, in the positive direction, to somewhere between –70 and –50 mV. When the sodium channel opens, it changes its conformation (Figure 13.6). After a few 10,000ths of a second, it again changes its conformation and closes, and the Na^+ ions cannot enter the nerve axon. The conformational changes of the sodium channel during depolarisation state are rapid, while those during the closing of the channel are delayed. The membrane potential starts recovering back to its resting stage as soon as the sodium channel is closed. The inactivated gate will not reopen

till the membrane potential reverses back to the resting negative potential. The potassium channel is closed at the resting stage preventing potassium ions flowing to the exterior. When the membrane potential moves in the positive direction, i.e. from –90 mV to zero then the potassium ion channel opens up and K^+ ions diffuse outward. As this process is relatively slow, it takes place when the sodium channel is inactivated. Thus the closing of the sodium channel and simultaneous opening of the potassium channel speeds up the repolarisation process.

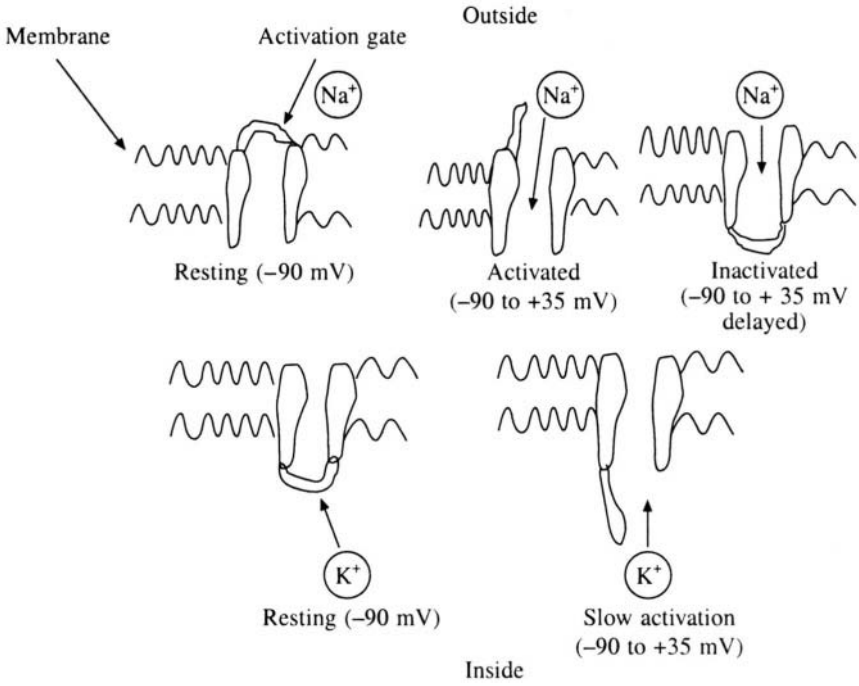


Fig. 13.6 Voltage gated sodium and potassium channels

13.3.2 Voltage clamp

Membrane potentials change very rapidly and it is possible to study the kinetics of these processes only when the membrane potential is held constant at a specified value, and if the measurements are carried out for a specified time. This is achieved by the voltage clamp technique. It was first successfully employed by Hodgkin and Huxley to measure the flow of ions through different channels. The voltage clamp has two electrodes inserted into the nerve fibre. One of these is for measuring the voltage and the other is to conduct electric current inside or outside the fibre (Figure 13.7). The investigator chooses a voltage to be established inside the nerve fibre, which is greater than the threshold value, and the voltage is clamped at the chosen value by a feedback system. The membrane potential (E_m) is measured by an intracellular electrode, and this is compared with the chosen voltage in a control amplifier. The output of this amplifier is connected to a second electrode, which passes a current into the cell, which is proportional to the deviation of the membrane potential from the desired voltage. Thus the deviation is compensated. It is to be noted that the compensating clamp current has the same amplitude but opposite polarity as the membrane current at the chosen potential. The current flow is measured by connecting the current electrode to an oscilloscope. For example, when the membrane potential is suddenly increased from –90 mV to zero by this method, then

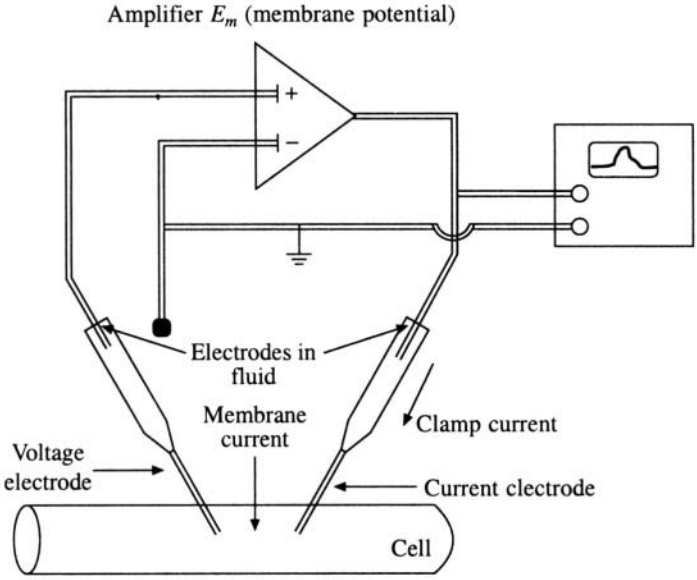


Fig. 13.7 Voltage clamp

sodium and potassium channels open up and the ions begin to flow through the channels. Electric current is injected through the current electrode to compensate for the net current flow through the channels and to maintain the intracellular voltage at zero (or a pre-determined value). The conductance, which varies with time, is measured. By selectively inhibiting one of the channels, the flow of ions through individual channels can be studied. Toxins like tetrodotoxin are used for blocking sodium channels while terra ethyl ammonium ion is used for blocking potassium channels. Figure 13.8 is an example of the results obtained from such measurements. Changes in conductance when the membrane

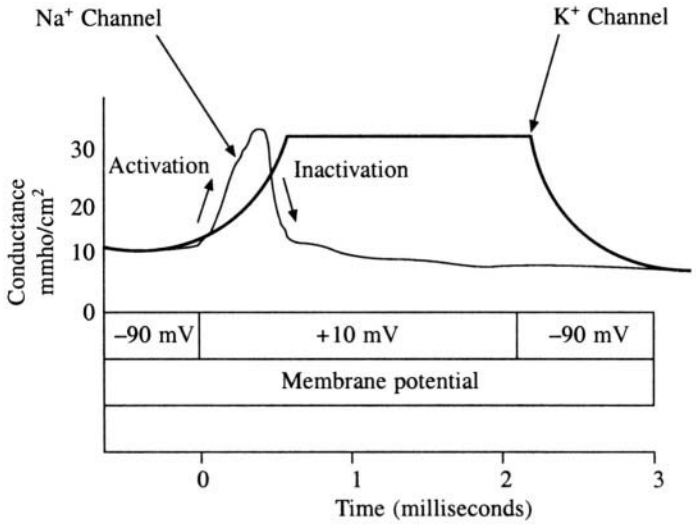


Fig. 13.8 Changes in conductance during membrane transport

potential is increased from its normal resting value of –90 mV to +10 mV for 2 milliseconds are shown. The sodium channels are activated and inactivated within this time while potassium channels are only activated during this time.

The behaviour of the nerve fibre in terms of changes in permeability of the membrane to Na and K ions has been formulated as a set of equations, known as the Hodgkin-Huxley equations, as follows:

$$g_K = n^4 \overline{g_K}$$

$$dn/dt = \alpha_n(1 - n) - \beta_n n$$

for K^+ conductance, where $\bar{g}_K$ is the constant maximum conductance. The state α is that in which the membrane allows K^+ to pass, while β is the state in which the membrane is impermeable to K^+ ions. n is the proportion of molecules in α state while $(1 - n)$ is the proportion of molecules in state. β_n is the reaction rate of the change from state α to state β and α_n is that of the change from β to α. At the resting stage β_n is large and α_n is small. Consequently most of the regulating molecules are in state β. On depolarisation α_n increases and β_n decreases depending on the membrane potential and this leads to a rise in n and hence in g_K. After about 10 milliseconds, α_n is larger than β_n and an equilibrium is reached where most of the molecules are in α state. n and g_K attain maximum values and this will last as long as the depolarisation state lasts. For sodium conductance, the Hodgkin-Huxley equations are as:

$$g_{Na} = m^3 \bar{g}_{Na}$$

$$dm/dt = \alpha_m(1 - m) - \beta_m m$$

$$dh/dt = \alpha_h(1 - h) - \beta_h h$$

$\bar{g}_{Na}$ is the constant maximum sodium conductance m and h denote the state of the two molecules which regulate sodium conductance. α_m, β_m, α_h and β_h are reaction rate constants that depend only on membrane potential but not on time. Though the conductance of sodium ions is similar to that of potassium, the equations will have to accommodate complications like inactivation of the sodium channel after about 1 millisecond. The variable '*h*' has been introduced to take care of the inactivation process.

13.4 Sensory Mechanisms—The Eye

An organism is continuously bombarded by different signals from the environment. AH the sense organs that are provided to the organism are transducers of energy; i.e. they convert energy in one form into another. Unlike neurons the response in these cases is not an 'all or none' process. The intensity of the response by the sensory organs will vary with the extent of the stimulation. Sensory detectors are classified into teleceptors, chemical receptors, somatic receptors and visceral receptors. Organs like eyes and ears are teleceptor as they receive information from distant objects. Sensors of smell and taste belong to the class of chemical receptors as they are excited by chemical stimulation. Touch, pressure, pain and temperature are experienced through the somatic receptors, while the information regarding body conditions like thirst, hunger etc are conveyed to the brain by the visceral receptors. When stimulated, each of these receptors will initiate nerve discharges in the form of action potentials. However, the way in which sensory information, such as the image on the retina or the pitch of a sound, is converted into the binary code of action potential is unknown. The action potentials generated by sensory nerves are usually evoked by synaptic stimulation and are therefore

called generator potentials. Generator potential magnitudes vary with the intensity of the stimulation of the receptors and this is achieved by modulating the frequency of a sequence of action potentials in such a way that it reflects the intensity of the stimulus. Let us look at the basic question of how a stimulus like light, sound or temperature is converted into electrical polarisation of the membrane in the case of visual receptors and auditory receptors.

13.4.1 The visual receptor

There are different kinds light sensitive receptors of which vary from the most primitive system like photokinesis (light induced motion) in plants to most elaborate ones like the vertebrate eye. A schematic diagram of the vertebrate eye is given in Figure 13.9. The lens system of the eye, which consists of cornea, aqueous humour, ciliary muscle, and vitreous humour of the eyeball helps in focussing the image of the object onto a layer containing visual receptors. The layer on which image is formed is called the retina. The iris is an important structure and it controls the amount of light entering the eye. A small iris opening increases the depth of focus in addition to restricting the distortions such as spherical aberration, field curvature etc. For example in the night, sensitivity is more important than depth of focus or acuity. Hence the iris diaphragm will be open to the maximum extent.

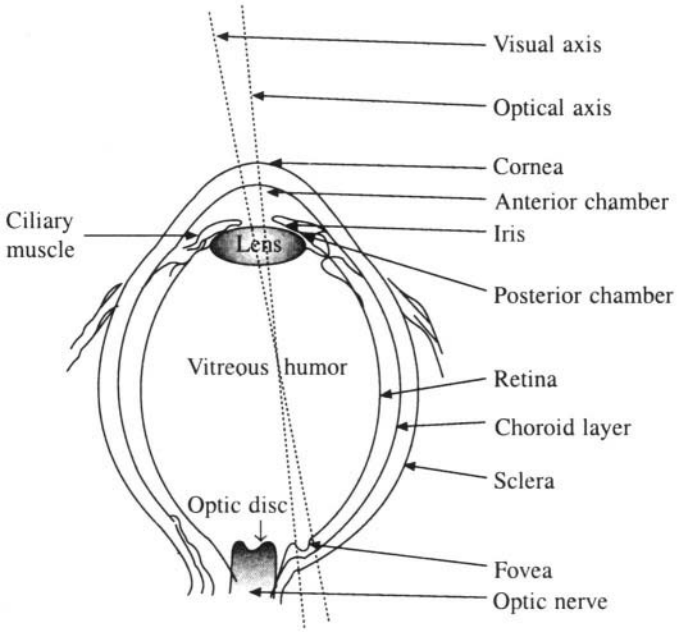


Fig. 13.9 Schematic diagram of the vertebrate eye

The crystallin lens or eye lens is the most important structure in the eye. The eye lens is a cellular structure, which is more curved at the rear than in front. The focus of the lens is varied by changing the curvature of the front face by the action of ciliary muscles. The space between the retina and the lens is filled with a jelly-like substance called vitreous humour. Vitreous humour is optically similar to the aqueous humour filling the cornea and the eye lens. Light enters the eye through the cornea and passes through the aqueous humour and the crystalline lens into the vitreous humour. The light is received by the retina and an image is formed. Figure 13.10(a) shows the geometrical optics of the

eye. Vision itself is made up of several processes. The photons (light) which fall on the retina trigger a series of events in the light sensitive photoreceptor cells. As a result the energy carried by the photons is transduced into chemical energy which in turn is transformed into an electrical signal. This signal is carried by the optic nerve to the brain for processing. The transduction of light into chemical energy is fairly well understood but the subsequent conversion into electrical signals is not completely understood. The image on the retina is formed by refraction of light by the cornea and lens and can be described by the well-known Snell's law

$$\sin \theta_2 / \sin \theta_1 = n_1 / n_2$$

where θ_1 is the angle made by the incident light with the normal and θ_2 the angle made by the refracted ray with the normal and n_1 and, n_2 are the refractive indices of the two media, respectively (Figure 13.10(b)). Before reaching the retina, light is refracted at three places. (1) At the surface of the cornea, going from the air into the cornea. (2) At the surface of the lens, going from aqueous humour into the lens. (3) At the posterior surface of the lens, in passing from lens to the vitreous humour, the refractive indices being $n_{air} = 1.0$, $n_{humour} = 1.33$, $n_{lens} = 1.413$. Thus the greatest refraction takes place at the posterior surface of the cornea. In order to calculate the image size we can replace the cornea-lens system by a single equivalent lens. In that case, we know that

Object size/Image size = Object distance/Image distance

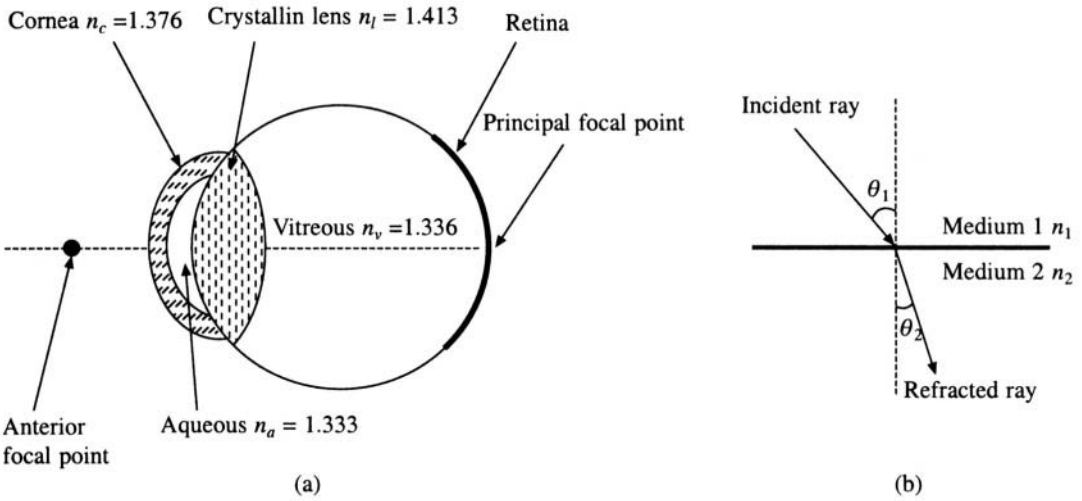


Fig. 13.10 (a) Optical properties of the eye. (b) Refraction of light

This expression can be used to calculate the image size. For example if the object distance is 1 km and the object size is 10 m then the image size = 0.15 mm, where the average image distance is taken to be 1.5 cm. The process of vision is dynamic, since the eye lens is not a rigid object with fixed focal length. The focal length can be adjusted by changing the curvature of the lens and this process is called accommodation. The standard relation between the object and image distances (p and q) is given by

$$1/p + 1/q = (n - 1)(1/R_1 + 1/R_2) = 1/f_L$$

where R_1 and R_2 are the front and rear radii of curvature of the lens, respectively, f_L is the focal length

of the lens, and n is the refractive index of the lens. If the focal length of the lens is fixed, then objects near the lens will be focussed behind the retina and vice versa. As the lenses have a flexible focal length the images are always focussed on the retina. The change of focal length is achieved by a change in curvature of the lens, which is made possible by the action of muscles attached to the lens. As age progresses the power of accommodation reduces and external lenses are used to correct this defect.

The image formed on the retina is received by photoreceptor which are called rods and cones (Figure 13.11). Cones are more concentrated at the central part of the retina while rods are more abundant at the periphery. Colour vision is due to cones while rods are responsible for vision in weak light. As rods are insensitive to colours, three types of cones have been implicated in colour vision. Though Newton's experiments demonstrated that a spectrum of white light consists of a continuum of colours, Maxwell showed that almost all colours could be obtained by mixing three primary colours. This suggests that at least three different receptors are necessary for colour vision. However, the perception of colour is achieved only by a great deal of processing of the signals by the brain. The fovea, which is the central part of the retina, has a high degree of acuity of vision and this is due to the exclusive presence of cones in this region. The optic disk lies very close to the fovea. This disk is also known as the blind spot. Objects focussed on this region cannot be seen as there are no cones or rods in the optic disk. The photoreceptor cells contain visual pigment molecules that are light sensitive. The light sensitive molecule in the vertebrate eye is rhodopsin and has a molecular weight

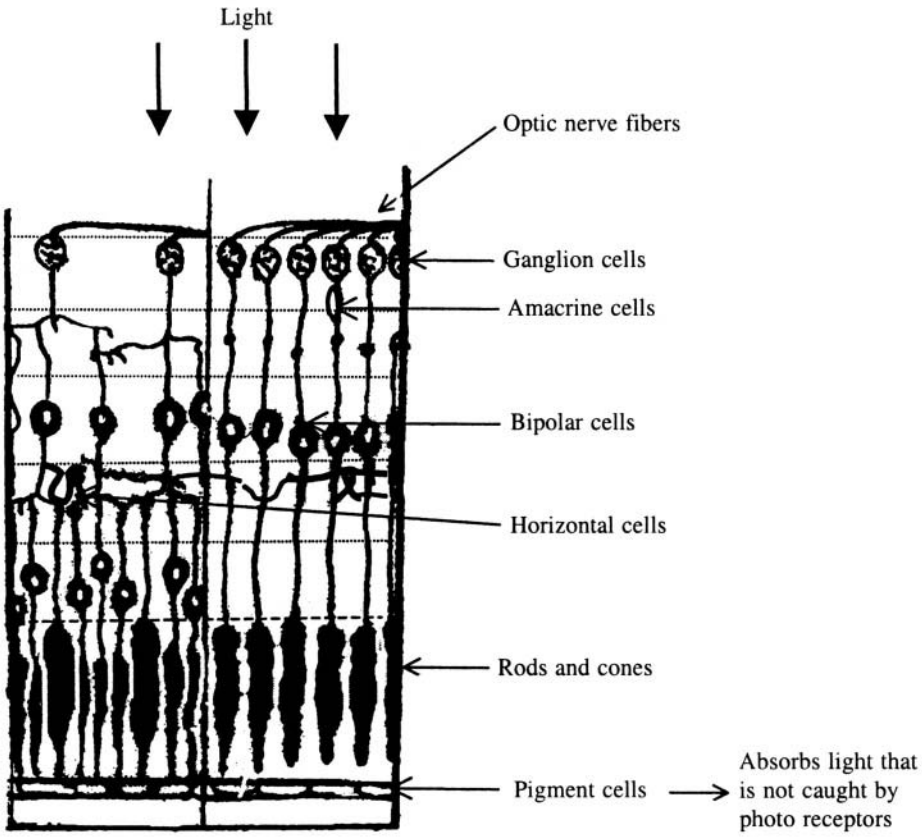


Fig. 13.11(a) Structure of retina

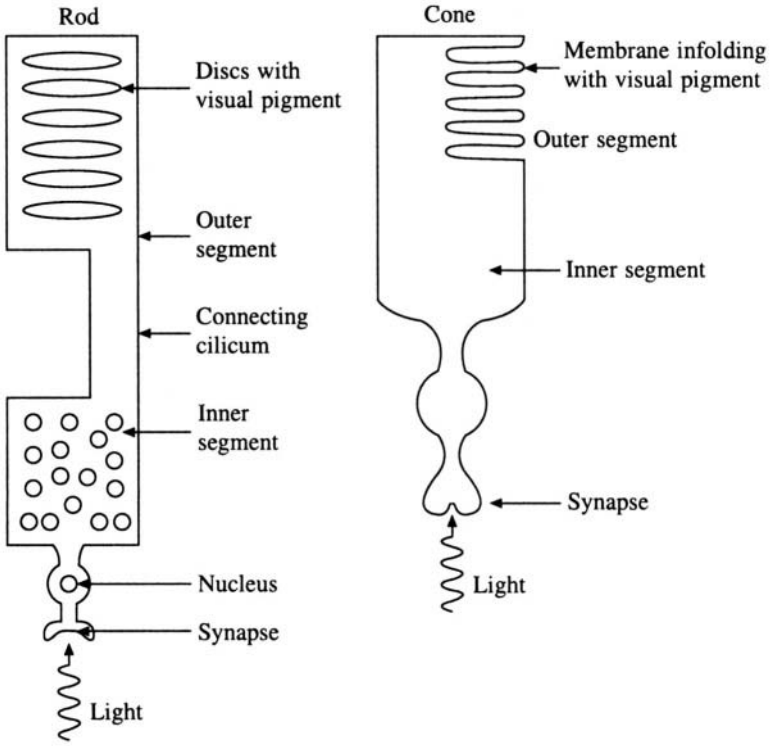


Fig. 13.11(b) Schematic diagram of rods and cones

of about 35 to 40 KDa. This consists of a single-polypeptide protein, opsin, to which other molecules are covalently linked, viz. a chromophore and two oligosaccharide chains. The chromophore is retinal and is responsible for absorption of light in the visible region. Retinal is a derivative of vitamin A and is obtained by oxidation of vitamin A by **NAD^+**. Figure 13.12 shows the two isomeric forms of retinal. Of the two oligosaccharides, at least one is composed of three *N*-acetyl glucosamine and three mannose units. Rhodopsin is a membrane protein and is insoluble in water. Figure 13.13 shows the structural arrangement of seven **α-helical** segments of rhodopsin. The absorption spectrum of rhodopsin is shown in Figure 13.14. The absorption maximum is at approximately 500 nm. This is also the most effective wavelength for human vision in a dark-adapted eye that has been exposed to weak light. In strong light the peak is shifted to 550 nm which is close to the absorption maximum of iodopsin. In strong light when the rod vision changes to cone vision, iodopsin is the visual pigment involved.

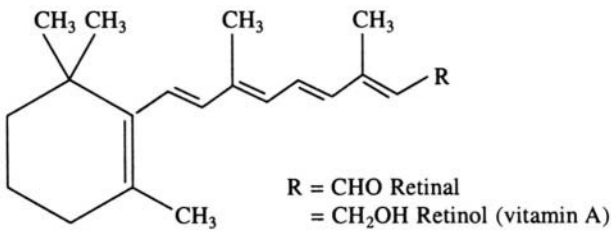


Fig. 13.12 Structure of retinal and retinol

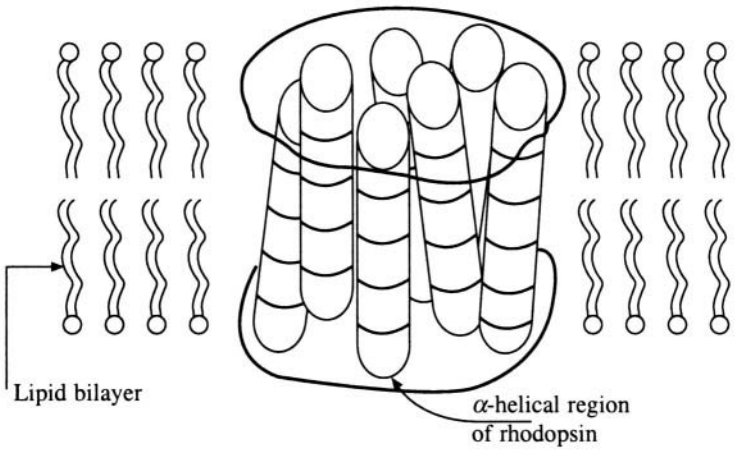


Fig. 13.13 Arrangement of the α-helical segments of rhodopsin in the lipid bilayer

The first step in the energy conversion procedure is the absorption of photons by chromophores. This causes a change in conformation of retinal from '11-cis' to 'all-trans' leading to detachment of the chromophore from opsin. This is known as the bleaching reaction. The recovery process involves reduction of retinal to 'all-trans' vitamin A, diffusion of vitamin A into the epithelium and re-conversion of vitamin A by light or enzymatically into the '11-cis' form. The '11-cis' vitamin A is then oxidised to '11-cis' retinal and linked to the protein opsin. The protein probably goes through a number of conformational changes before the 'all-trans' retinal comes off. Thus detachment and reattachment of the chromophore is the crucial part of the photochemistry of vision. An enzyme in the retina maintains an equilibrium between retinal and retinol (vitamin A). Generally a higher concentration of the alcohol (retinol) is maintained. However if the aldehyde (retinal) concentration is low then it

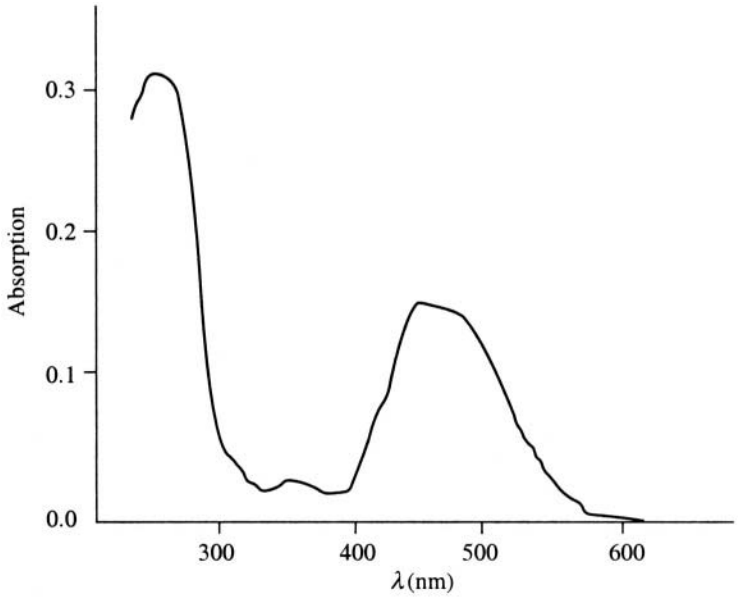


Fig. 13.14 Absorption spectrum of rhodopsin

is produced from the alcohol by the action of an oxidoreductase enzyme and NADPH. The vitamin produced by oxidoreductase enters the blood through the rod membrane and is maintained at a constant concentration by the liver (Figure 13.15).

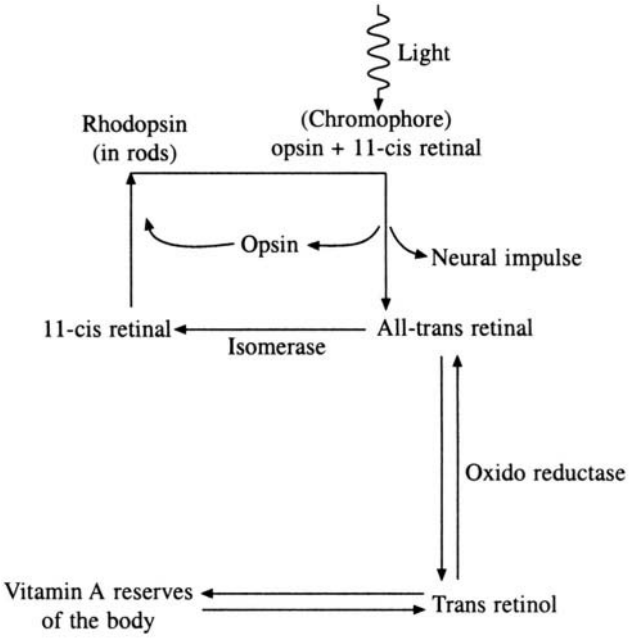


Fig. 13.15 Visual cycle undergone by rhodopsin

13.4.2 Electrical activity and visual generator potentials

When the rhodopsin molecule absorbs a photon, it undergoes a change in its electrical state, which is passed on as a membrane potential. This electrical signal, which is known as receptor potential, is proportional to the intensity of the light stimulus. The receptor potential is processed at the synapse and translated as series of nerve impulses, which are carried to the brain for visual perception. Information about the firing of action potentials in optic nerve fibres has been obtained by implanting microelectrodes in selected places. In vertebrates the ionic permeability decreases while in invertebrates the reverse occurs. The measurements show that the K^+ ion concentration is about 20 times more inside the cell than outside while the sodium ion concentration is more on the outside than in the inside. This is the same as the situation seen in nerve cells. The association of electrical impulses with the visual process is different from that of transmission of electrical impulse in the axon. The receptor cells behave like photocells. The membrane potential of the vertebrate photoreceptor is also mainly a diffusion potential. The potential difference across the cell membrane in the inner segment is larger than in the outer segment and this causes a permanent ion current in darkness. This current flows from the inner segment to the outer segment through the extracellular medium. In the dark, the external membrane of the rod outer segment is more permeable to Na^+ ions while the membrane of the inner segment is preferentially permeable to potassium ions. On illumination, the permeability for the sodium ions across the outer segment cell membrane drops and this drop is greater if the intensity is higher. This transient light induced reduction of conductance causes a reduction of the membrane current and a transient increase in the membrane potential (i.e. the membrane is more polarised) and

this is communicated to the horizontal cells. The signal is actually produced in the amacrine and ganglion cells, and is transmitted to the brain by the optic nerve.

13.4.3 Optical defects of the eye

A person with normal vision can see objects as close as 250 mm. In order to focus objects from 250 mm to infinity the crystalline lens curvature has to be changed by about 20%. The strength of the human eye is about 48 diopters. If the eye is stronger than this, then images of distant object will be focussed in front of the retina. Thus distant vision will be not be clear as compared to near vision and the person is said to be myopic or suffering from myopia. This can be corrected by using a diverging lens in front of the eye (Figure 13.16). If the strength of the eye is too weak, then the images of near objects will be formed behind the retina. This condition is known as far-sightedness or hypermeteropia, and requires a convergent lens for correction. The adaptability of the ciliary muscles drastically reduces, as the age increases. This also leads to long sight and a person who is myopic would require bifocals for correcting the near and far vision independently. Astigmatism is another eye defect due to the deviation of cornea from sphericity. When astigmatism is present, a point object does not form a point image on the retina. This is because the eye lens has different focal lengths for different directions. This defect can be corrected by using cylindrical lens.

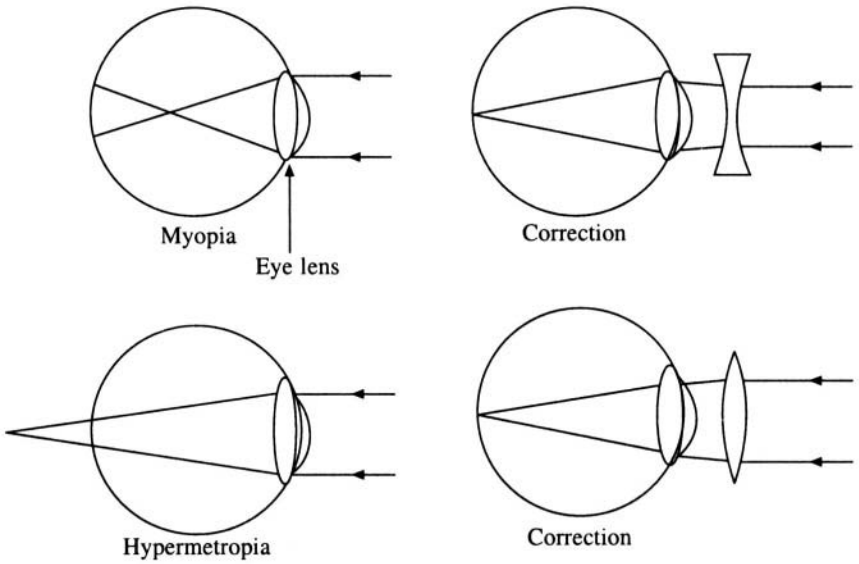


Fig. 13.16 Defects in vision and their corrections

The response of the eye extends beyond the visible spectrum. For example the cornea absorbs ultraviolet light which can damage it. Hence dark glasses are recommended in strong sunlight.

13.4.4 Neural aspects of vision

The retina acts as a photo-neural transducer i.e. it converts light energy into a neural response. This response, which is in the form of a spike potential, travels along the optic nerve and reaches the central nervous system (CNS). From the central nervous system, the neural response reaches specific areas in the cerebral cortex. The information is analysed both in the retina and in the CNS to generate the sensations of shape, colour, brightness, etc. The responses from either eye for an object are

projected on to the occipital lobe of the cerebral cortex opposite to the half of the visual field containing the object. There are three layers of cells which are important in the function, viz. the retinal cells, intermediate cell and a third layer of cells. These layers are interconnected and the information flows from retinal cells to the brain through these layers of cells and optic nerve. The information coming out of the optic nerve is divided into two bundles before reaching the brain. The visual cortex in the brain is responsible for the analysis of these impulses and subsequent vision. There are a few fibres that go to the midbrain rather than to visual cortex directly. Information such as the intensity of light, clarity of vision, etc., are processed in the midbrain and a feedback is sent to the appropriate regions of the eye. Figure 13.17 shows the neural connections between the eye and the visual cortex. From this figure one can see that the left side of the brain receives information from the right side of the visual field and the right side of the brain receives signals from left side of the visual field. They run through the optic chiasma and the information from both the eyes are put together to achieve binocular vision. Every point in the retina seems to have a corresponding point in

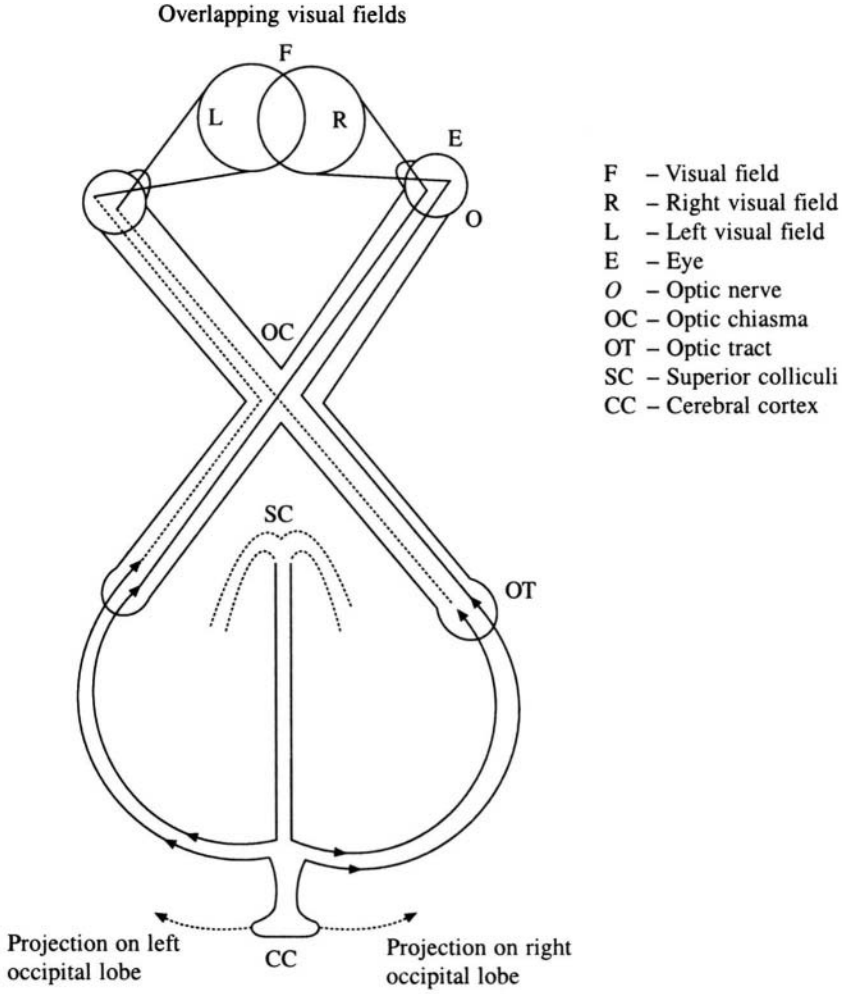


Fig. 13.17 Pathways of vision in the brain

the visual cortex. The points that are close together on the surface of the retina are also close together in the visual cortex. However, the central field of vision, which occupies a very small region on the retina, is expanded and is represented by several cells in the visual cortex.

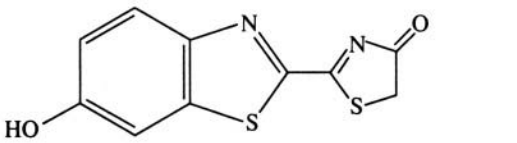
13.4.5 Visual communications, bioluminescence

Several insect families communicate with others in their species by producing a light, without accompanying heat. The visual signals can vary in intensity, colour and periodicity. The production of light without heat by living organisms is called luminescence. Most bioluminescent signals are not continuous and are mostly produced in the night or in special habitats like caves, ravines etc. A diverse variety of organisms like bacteria, fungi, sponges, corals, and fish can produce luminescence. Amphibians, reptiles, birds and mammals do not possess this property of producing light. The bioluminescent reaction in fireflies has been well characterised. The two substances responsible for this reaction, viz. luciferin and luciferase, have been isolated and crystallised, and their structures have been established. Figure 13.18 gives the chemical structure of luciferin. The reaction involving these two chemicals is a luciferase-catalysed oxidation using molecular oxygen and a substrate, viz. D(–)luciferin. A large amount of energy is liberated in this reaction and the intermediate or product is left in the excited state. When the excited product returns to the ground state, light is emitted. ATP is hydrolysed in this reaction, which is written as follows:

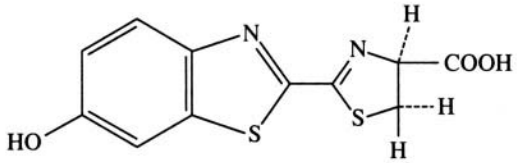
$$\text{Enzyme} + \text{D(–)luciferin}(LH_2) + \text{ATP} \overset{Mg}{\Leftrightarrow} \text{Enzyme–}LH_2\text{–AMP} + PP_i$$

$$\text{Enzyme–}LH_2\text{–AMP} + O_2 \rightarrow \text{Oxyluciferin} + \text{AMP} + CO_2 + \text{Light}$$

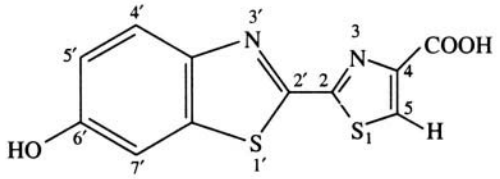
where **enzyme–LH_2–AMP** is the enzyme bound-luciferyl adenylate.



Oxyluciferin



D(-) Luciferin



Dehydroluciferin

Fig. 13.18 Chemical structure of luciferin

13.5 Physical Aspects of Hearing

Hearing, like vision, is an important sensory mechanism that enables many animals, including humans, to perceive their environment. Though the biophysics of hearing has been one of the earliest to be studied, information on the transduction of sound is meagre. This is due to the fact that the sensory organ, viz. the ear, will have to be intact for it to function and it is very difficult to perform experiments and measurements on the inner ear, where transduction takes place. However the structure of the ear is known in very great detail from anatomical and histological studies, and this has helped to reconstruct the hearing process.

Hearing is actually the perception of pressure waves that are generated by a vibrating medium. The ear, which is the hearing apparatus, is homologous in most species though their sensitivity varies. The frequency range of the normal human ear is 20 to 20,000 Hz.

13.5.1 The ear

The ear is divided into three parts - the outer, the middle and the inner ear. Figure 13.19 shows the human ear. The outer and middle ear are rilled with air and their primary function is to conduct the sound waves to the inner ear, where the acoustic energy is converted to neural spikes, which are analysed by the brain. The outer ear consists of an external auricle, a narrow tube called the ear canal or external auditory meatus and the ear drum or tympanic membrane. The ear canal is similar to a closed end organ pipe with the tympanum serving as the end closing membrane. The vibration of air molecules produced by an external sound is passed on to the small bones in the middle ear. The middle ear has three small bones or ossicles called malleus, incus and stapes. The ossicles amplify the acoustic pressure of vibrations transmitted via the tympanum. They also restrict the amount of energy that is transmitted at high sound levels. Thus when a high sound reaches the system then the amplification at the middle ear is decreased and this acts as a volume control. The ossicles also discriminate against

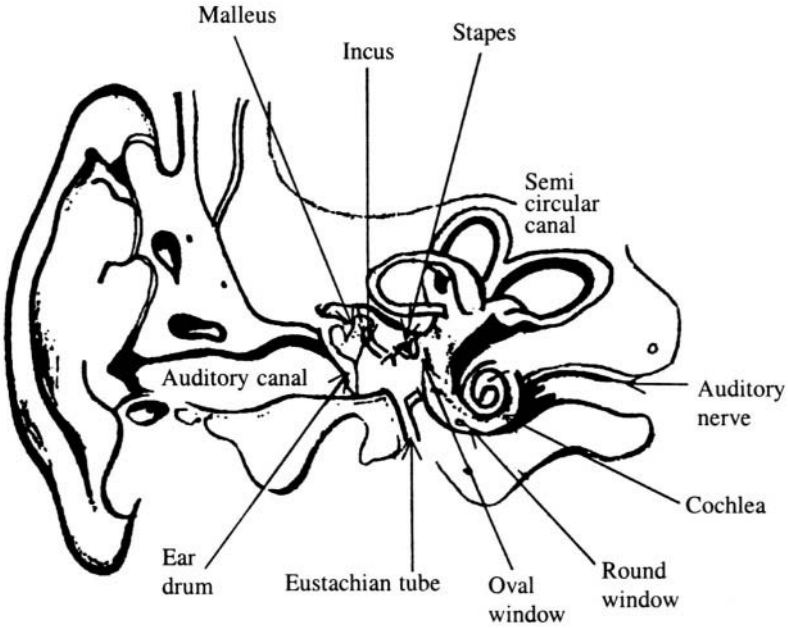


Fig. 13.19 Human ear

the vibrations reaching them through the skull. The outermost ossicle or the malleus presses against the tympanic membrane while the innermost one, the stapes pushes against a membrane called the oval window. The oval window separates the air-filled middle ear from the liquid-filled channels of the inner ear. There is a second membrane known as the round window, which separates another channel of the inner ear from the middle ear. The maximum pressure amplification achieved by the outer and middle ear together is about 35 dB. The middle ear is connected to the pharynx through a tube called the eustachian tube. The eustachian tube prevents distortion in the shape and position of the tympanic membrane when a change in pressure takes place due to atmospheric conditions or changes in the altitude. The mammalian inner ear consists of several portions of which only the cochlear portion is involved in hearing. The cochlea has two and half turns and has three parallel fluid-filled tubes. These tubes are called tympanic, vestibular and cochlear ducts. The individual ducts or canals are separated by membranes. Reisner's membrane separates the cochlear duct and the vestibular duct while the basilar membrane separates the tympanic duct from cochlear duct (Figure 13.20). The organ of corti, which is situated in the basilar membrane, is considered as a neuro-mechanical transducer. This organ is a complicated system of cells. It includes hair cells. The bending of these hair cells excites the nerve endings located in the organ of corti. These excitations are transmitted to the brain where it is perceived as sound. In order to understand the physical aspects of hearing a basic knowledge in acoustics is required.

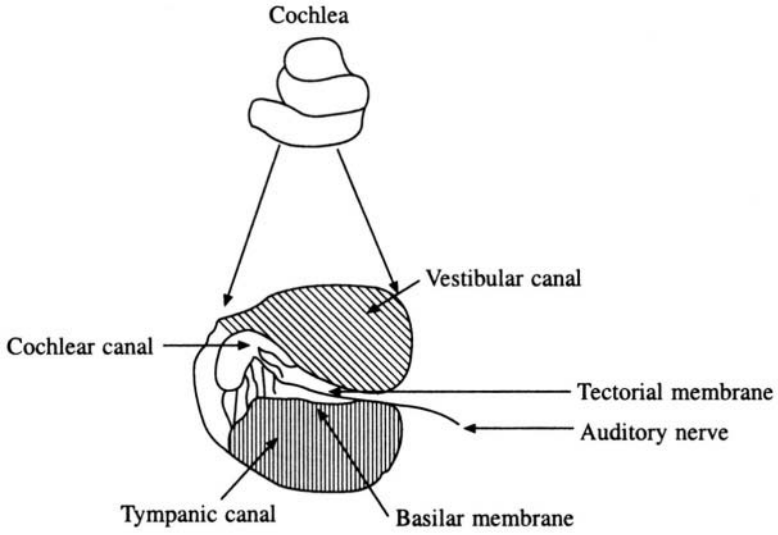


Fig. 13.20 Cochela and its cross section

13.5.2 Elementary acoustics

Sound is actually an elastic disturbance in a continuous medium. Supposing a particle in the medium is disturbed from its equilibrium position by a sound wave. If v is the frequency of the displacement, and ψ is the amplitude, then we can write the expression describing this effect as

$$\psi = A \sin(2\pi\nu t) + B \cos(2\pi\nu t) \text{ or}$$

$$\psi = Ce^{2\pi i\nu t}$$

where the amplitude varies as a function of time t. Most sound waves are mixtures of frequencies. If

the velocity of sound waves is c then $\lambda = c/\nu =$ **wavelength.** In a gas the pressure variation produced by the passage of sound wave is very rapid and hence the process is assumed to be adiabatic, i.e. occurring at a constant temperature. The process of hearing can be considered as the conversion of mechanical energy in the form of pressure variations in the air to electrical energy in the form of nerve pulses, which are transmitted to the brain for processing by auditory nerves. The conversion starts at the tympanic membrane in response to the pressure variations in the atmosphere and causes the three bones in middle ear to deflect. The three small bones in the middle ear have a mechanical linkage, and cause compressions in the cochlear fluid. The inward motion of the stapes causes downward motion of the basilar membrane, which triggers the nerve signal. The movement of the basilar membrane is a travelling wave, with its maximum amplitude at some small distance from the oval window. The responses of basilar membrane to sound have been measured using microelectrodes placed in the cochlear space and in the fibres of auditory nerves. Some typical responses are and that the displacements are inversely proportional to the frequency of the exciting sound wave, are shown

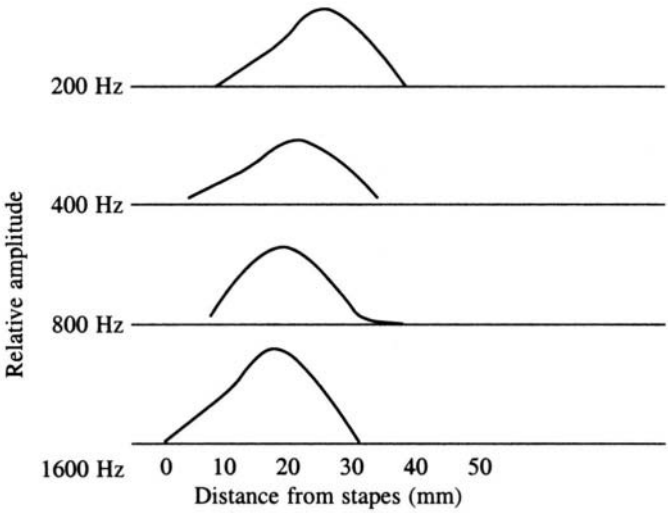


Fig. 13.21 Amplitude of the basilar membrane displacements

in Figure 13.21. The magnitude of amplification is estimated as follows. The force on the stapes is

$$F_{\text{stapes}} = 1.3 \times 52 \times P_{\text{tympanum}}$$

where 1.3 is a constant signifying the mechanical advantage of ossicles, $52\ \text{mm}^2$ is the area of contact between the membrane and the hammer, and P is the pressure. The area of contact between stapes and round window is $3.2\ \text{mm}^2$. Therefore the pressure at the window is

$$P_{\text{window}} = F_{\text{stapes}}/3.2 \text{ (since Pressure = Force/Unit area)}$$

Therefore

$$P_{\text{window}}/P_{\text{tympanum}} = 68/3.2 \cong 21$$

This is roughly the maximum amplification observed. In humans, the average value of magnification is about 17 and is less than the calculated value because of friction, which has not been taken into account in the above calculations.

13.5.3 Theories of hearing

The different theories that have been formulated to explain hearing are known as the spatial and temporal theories. The spatial theory proposed by Helmholtz suggests that a particular point on the basilar membrane responds to a particular audible frequency. Anatomical investigations have shown the basilar membrane to have a fibrous structure and individual fibres are expected to resonate at a particular frequency. The Helmholtz theory met with a few difficulties. Experiments conducted with resonators showed that the fibres respond to a narrow band of frequencies rather than a single frequency. Secondly, any resonating system produces an 'echo' which continues even after the excitation of resonance has stopped. However, no such effect is observed in the hearing process. Nevertheless, hearing loss due to damage to particular regions of the membrane clearly supports the view that the loss is due to the lack of response by the fibres expected to resonate at those frequencies. Helmholtz compared the resonance of the fibres to that of a stretched string. The resonating frequency of such a stretched string is given by

$$\nu_{res} = {}^1/_2\, l\sqrt{(T_0/\rho)}$$

where l is the length of the string, T_0 is the tension on the string, and ρ is its density. Using the above equation we can calculate the frequencies over which the ear is sensitive. When this is done we find that the audible frequency range is reduced by five octaves from the experimentally known values. This shows that Helmholtz's theory does not completely represent the hearing process.

Bekesey showed that in addition to the variable compressions one should take into account the hydrodynamic waves that are produced in the cochlear fluid in explaining the hearing process. He could successfully predict the point of maximum response in the basilar membrane as a function of frequency. The maximum displacement position is dependent on the initial frequency. The mathematical theory dealing with hydrodynamic waves is quite complex and a simplified explanation is given here. We have seen earlier that the vibration of the stapes produces a wave in the cochlear fluid, which spreads from the oval window throughout the cochlea in about 25 microseconds. This leads to vibrations in the basilar membrane, but the membrane being less elastic at the base, the displacement moves towards the apex. Thus the fluid and the membrane form a coupled system and hence individual parts cannot resonate separately. As mentioned earlier, the position of the maximum displacement depends on the initial frequency; for example when the frequency is low the entire membrane moves at the tip but if the frequency is high then the oscillations are produced near the oval window. This explains why damage to certain regions of the membrane leads to hearing loss at certain frequencies. Thus, the high frequency response depends on the functional response at the base while low frequency depends on the condition of the apex. The organ of corti acts as a transducer and the motion of the basilar membrane is converted into an electrical signal and propagated along the nerve fibres to the auditory cortex of the brain. The threshold potential required to fire a cell depends on the frequency of stimulation. The processing at auditory cortex of the brain is still an unsolved problem.

13.6 Signal Transduction

Biologically significant information is generally transmitted through chemical signals. The message carriers are special chemicals. Messages are also transmitted by intracellular substances, like mRNA, cAMP (cyclic adenosine monophosphate), etc. Communications between cells and organs are carried out through substances like hormones. Neural communication, as seen earlier, is through neurotransmitters. Chemical signals like odours emanating from bugs, skunks, flowers etc., are useful in communication between organisms of the same species or different species. The antigens that stimulate the production of antibodies can also be considered as biological signals.

In any signal mechanism there is a sender, a receiver and mode of transport of signal. In biological systems the receiver of chemical signals are chemosensitive organs, chemoreceptor cells and other receptor molecules. For the recognition of the signal, the receiver molecule must have chemical specificity. The nature of the response depends on the concentration and nature of the signal substance. In most cases the chemical specificity is due to macromolecules which are the receptors. The molecules on the cell surface that enable cell-cell recognition (e.g. glycoproteins, glycolipids) can also be considered as chemical signals. The three stages in molecular recognition are: (a) capturing of the signal substance by the receptor molecules, (b) causing of the physiological effects and (c) removal of the signal substance by decomposition or by removal from the receptor.

13.6.1 Mode of transport

Transport of signals could be due to diffusion or convection. The diffusion process is effective only when the distance between receiver and sender is small (of the order of a few tens of nanometres) and the concentration of signal substance is high. For example, in nerve transmission the neurotransmitters travel only a distance of about 20 nm. In the case of hormones, which may travel from one organ to the other, the signal substance is first adsorbed on the cell membrane, and then diffused through the membrane for action. If the affinity for the signal substance to the receptor molecules on the membrane is high, then these molecules could accumulate specifically on this cell, and help in transporting the signal. Thus the strong affinity of the lipophilic surface of the olfactory organ for odour-producing substances helps insects to respond to even low concentrations of the signal substances in the air.

Convection is another method of transport of signal molecules and is especially useful when the distance between the sender and the receiver is large. The time taken for transport by diffusion is

$$t = x^2/D$$

where x is the distance to be travelled, and D diffusion coefficient. Hence the time taken for transport by diffusion over large distances is large. Convection plays an important role under these circumstances. For example odorous substances are known to move across distances of even several kilometres and yet be recognised at the destination.

13.6.2 Signal transduction in the cell

The chemical signals are recognised by receptor proteins. The binding of the signal molecules to the active site of the receptor is reversible, due to its non-covalent nature. The strength of the binding is also dependent on the extent of complementarity between the signal molecule and the active site of the receptor. The receptor active site could be, for example, the opening of an ion channel in the membrane of the muscle end plate when the neurotransmitter acetylcholine binds to the receptor molecule. Another such example is the binding of antigen molecule to the hypervariable region of antibodies. In the case of insect olfactory cells, even a single odour molecule is sufficient to trigger a nerve impulse. This is made possible by opening of a single ion channel in the sensory cell membrane.

If the signal substance is continuously produced, then accumulation of the signal substance in the transporting medium can take place. This has to be avoided by removal of the signal substance by decomposition, inactivation or by dilution due to diffusion. For example, in the antigen-antibody complex, irreversible chemical binding inactivates the signal substance. In the case of neurotransmitters the acetylcholine molecule is broken down very rapidly by acetylcholine esterase. The receptor surface of the olfactory cells is cleaned by being continually rinsed by convection and by the flow of fresh mucosa over the cells.

14

Origin and Evolution of Life

14.1 Introduction

Life can be broadly defined as something that can replicate and sustain itself. No other precise or exact definition of life is available. Biochemists and molecular biologists generally describe the properties of living organisms as the following.

1. Systems that are separated from the environment by an envelope (membrane).
2. Systems that have basic structures and mechanisms common to all. For example, the genetic code or the mechanism of translation of the genetic messages into polypeptides is essentially the same in all.
3. Systems that sustain themselves by an internally controlled metabolism, which is essentially the exchange of matter with the environment and the synthesis of energy rich compounds.

It is possible to find living systems that do not possess all three properties listed above. But certainly every system that can be called 'Life' will possess at least one or two of them.

Evolution can be defined as change or adaptation by the organism, in response to environmental factors. Evolution can be classified into chemical evolution, molecular evolution and Darwinian evolution. Chemical evolution is generally used to denote the prebiotic (before biology) formation of organic molecules in the period between the formation of earth and the first appearance of living cell. Molecular evolution refers to the changes in living systems that have occurred thereafter, leading eventually to the formation of unicellular and multicellular organisms. Darwinian evolution refers to the events thereafter, which have lead to the species seen today on earth. Figure 14.1 provides the evolutionary sequence schematically.

The origin of life is linked to the environmental conditions prevailing at that time and also to the elements available. Any attempts to propose a theory for origin of life should involve a chain of logical steps that are consistent with known laws of physics and chemistry. The interesting possibilities

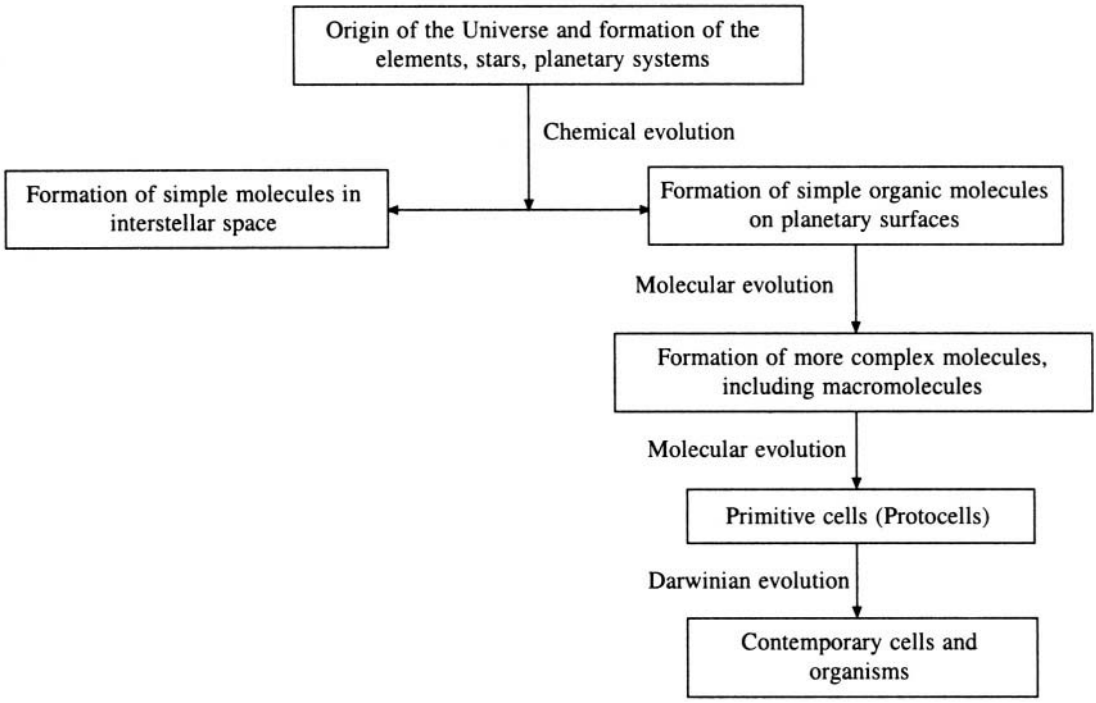

Origin of the Universe and formation of the elements, stars, planetary systems
Chemical evolution
Formation of simple molecules in interstellar space
Formation of simple organic molecules on planetary surfaces
Molecular evolution
Formation of more complex molecules, including macromolecules
Molecular evolution
Primitive cells (Protocells)
Darwinian evolution
Contemporary cells and organisms

Fig. 14.1 Scheme of evolution

that arise in the discussion of origin of life include the following: (1) Life evolved elsewhere and was introduced on the earth. (2) Life as we know it today, originated fully developed, all-at-once, spontaneously by chance. (3) Life originated on earth first in a primitive form and then evolved gradually to the present-day forms. It is difficult to prove that life originated elsewhere and was later implanted on the earth. Similarly it is hard to expect the present day system to have originated spontaneously, i.e. to expect very high order to emerge from total disorder by chance. Thus we are left with the third option. To explore this option further, we have to consider how primitive molecules could have arisen on earth in the environment which prevailed at those early times, soon after the cloud of gases condensed into a solid planetary mass with continents and oceans. It becomes necessary to know the conditions prevailing at that time on earth.

14.2 Prebiotic Earth

The earth is believed to be 4500 to 4750 million years old and it has been suggested from fossil records that life in the form of unicellular organisms appeared about 3000 million years ago. The atomosphere of the primitive earth was devoid of molecular oxygen, which made it more suitable for chemical evolution. UV light could reach the earth undiminished due to the absence of an ozone layer. Other forms of energy that were available include ionising radiation, electric discharges (thunderstorms), heat etc. The most important organic elements like carbon, hydrogen, nitrogen and oxygen were also the most abundant elements on earth, besides helium, neon and silicon. The synthesis of carbon based compounds would have proceeded at fairly high rates in such an atmosphere. This is borne out by the fact that most of the interstellar compounds are carbon based, like **CO, CH_4,**

CH_3OH, cyanoacetylene etc. Also carbon compounds are more stable to the action of water, oxygen, UV light and heat.

It is very difficult to know exactly how life originated on eath. We can reconstruct the earliest stages of evolution only by conducting experiements in the laboratory under the conditions knowns to be prevalent at that time or by looking for evidence in fossils. Fossils are the remains of living matter after their slow mineralisation into stone. These fossils can only give information about the gross structure of the cell, as most of the organic matter would have been destroyed by repeated heating and cooling of the rocks due to change in the environment. However, molecular fossils, which is the name given to those structures or functions in the cell that are shared by diverse present-day organisms, may yield clues to build a coherent picture of primordial events.

14.3 Theories of Origin and Evolution of Life

The evolution of life after the formation of the first cell can be considered to have been controlled by fairly well established Darwinian principles. However, prebiotic evolution is an area of intense debate as it depends on the interpretation of fossil remains or on extrapolations from existing life forms. There are no fossils of the early period of evolution i.e. 4 billion years ago. Whenever experimental evidence is lacking, extrapolations have been made or theories have been proposed. Not surprisingly, these contain several contradictions.

Eigen proposed a theory for the self-organisation of macromolecules and the emergence from chaos of ordered structure, which in turn created information. He said that this is possible only when we have an autocatalytic open system, which is far away from equilibrium. Accidental ordering from chaos has a very low probability. For example, given that a polynucleotide chain of 100 nucleotides is to be formed from four nucleotides, the number of possible sequences is equal to 4^{100} or 10^{60}. For any particular primary structure to evolve, the probability is close to zero. Eigen proposed a model in which he assumed that macromolecules do not make an exact copy of themselves, (i.e. do not self-reproduce) but replicate by synthesising complementary strands. In a more complex system of reactions, there could be a cycle of molecules and reactions, each supporting the next in line. Such a cycle of reactions is called a hypercycle (Figure 14.2). Eigen's hypercycles have the following properties: (1) All species (molecules) linked through a cycle have a stable and controlled coexistence. For example, a hypercycle can be formed by a sequence of reactions in which the product of one reaction may become the reactant of the next reaction and so on. The citric acid cycle is one such set of reactions. (2) A hypercycle can increase or decrease in size if these changes have a selective advantage. (3) Once established, a hypercycle has a distinct advantage over a new, yet-to-be-established species (of hypercycle). To explain how these hypercycles come into existence in the first place, Eigen proposed that the first self-reproducing system with enough information content should have been a short tRNA-type molecule. This lead to the proposal of a so-called 'RNA world', in which the catalytic molecules as well as the information storage and transfer molecules are RNA. Eigen and Schuster have extended this theory with models for the origin of the genetic code and the translation system.

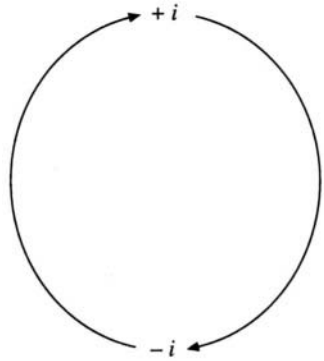


Fig. 14.2 Eigen's hypercycle

Kuhn's model proposes the formation of short polynucleotide chain like pre-tRNA molecules that have secondary and tertiary structure. The periodic changes in the environment act as a selection factor. Molecules are assumed to have been synthesised in the pores of clay minerals. This theory also proposes steps related to self-assembly of tRNA molecules, synthesis of proteins, formation of membranes, the genetic code etc.

Ebeling and Feistel's theory assumes that compartmentation of the reactants and products started in the early stages of chemical evolution. This compartmentation, also known as coacervate formation, gave rise to micro-reactors, which played the role of primitive cells during early evolution.

14.4 Laboratory Experiments on the Formation of Small Molecules

At the molecular level all forms of life appear to be similar and hence it is not inappropriate to assume that life on earth descended from single ancestral life form. Chemists have shown that essential building blocks of life can be synthesised from elements that were present on the primitive earth. This is also known as prebiotic synthesis. Miller's experiment showed for the first time that simple amino acids could be assembled spontaneously from the constituents of primitive earth.

14.4.1 Prebiotic synthesis of amino acids—Miller and Urey's experiment

The experiment attempted to simulate the pre-biotic atmosphere of the early earth. Air was pumped out of a small flask and then a mixture of ammonia, methane and hydrogen was added. Water vapour was introduced by adding water and boiling it. This also helped to mix and circulate the gases into a larger flask that was connected to the smaller one. An electric spark was generated across a spark gap set into the arrangement (Figure 14.3). The product of the discharge were condensed by a condenser and washed through a U-tube into the small flask. The non-volatile products remained in the small flask, but the volatile products re-circulated past the spark. The reduced forms of carbon, nitrogen and oxygen in the gas phase represented the prebiotic atmosphere; the liquid phase represented the ocean. When the products were analysed, it was seen that many different organic molecules had been synthesised. The results of Miller and Urey's experiment were unexpected in two respects. A small number of relatively simple compounds accounted for a large proportion of the products. Further the major products themselves were not a random selection of organic compounds but included a surprising number of substances that occur in living organisms (Table 14.1).

Table 14.1 Compounds that were synthesised in the Miller-Urey experiment

Compound	Yield (%)
Glycine	2.1
Sarcosine	0.25
α-Alanine	1.7
α-amino isobutyric acid	0.007
Aspartic acid	0.024
Formic acid	4.0
Urea	0.034
β-Alanine	0.76

α-Alanine was produced in greater quantity than **β-alanine.** When the experiment was performed over a longer period, large amounts of cyanide and aldehydes were formed during the first 125 hours

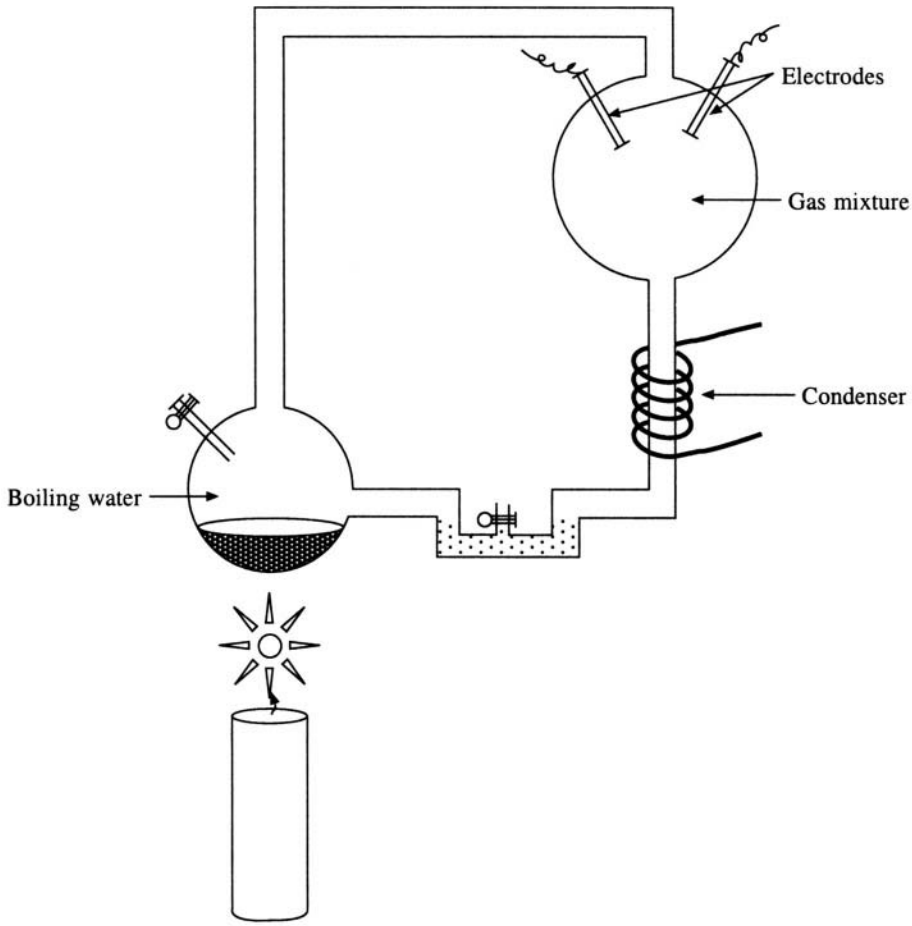


Fig. 14.3 Miller and Urey's experimental setup

of sparking and then the rate of their synthesis fell off, whereas the amino acids were formed at a more or less constant rate throughout the run. The mechanism proposed for the formation of glycine is as follows:

$$CH_4 + H_2O \rightarrow H_2CO + 2H_2$$

$$CH_4 + NH_3 \rightarrow HCN + 3H_2$$

$$H_2CO + HCN + NH_3 \rightarrow CH_2(NH_2)\text{–}C \equiv N + H_2O$$

$$CH_2(NH_2)\text{–}C \equiv N + 2H_2O \rightarrow CH_2(NH_2)\text{–}COOH + NH_3$$

The general reaction can be written as

$$RCHO + HCN + NH_3 \rightarrow \underset{\displaystyle NH_2}{\overset{\displaystyle |}{RCH}}\text{–}C{\equiv}N + H_2O$$

$$\underset{\displaystyle NH_2}{\overset{\displaystyle |}{RCH}}\text{–}C{\equiv}N + 2H_2O \rightarrow \underset{\displaystyle NH_2}{\overset{\displaystyle |}{RCH}}\text{–}COOH + NH_3$$

The above reactions depend on the formation of hydrogen cyanide.

14.4.2 Prebiotic synthesis of purines, pyrimidines and nucleosides

The most important purines are adenine, guanine and, to a lesser extent hypoxanthine (Figure 14.4). If concentrated ammonia solution containing 1-10 M cyanide is refluxed, adenine is formed in yields up to 0.5 % along with large amounts of the usual hydrogen cyanide polymer. This reaction does not easily explain the synthesis of purines under prebiotic conditions since useful yields of adenine cannot be obtained except in the presence of 1.0 M or stronger ammonia. The highest reasonable concentration of ammonia or ammonium ion that can be postulated in oceans and lakes on the primitive earth is about 0.01 M. In order to accommodate this fact, an alternative mechanism has been proposed. This mechanism makes use of an unusual photochemical rearrangement, which does not involve additional reagents as shown in Figure 14.5.

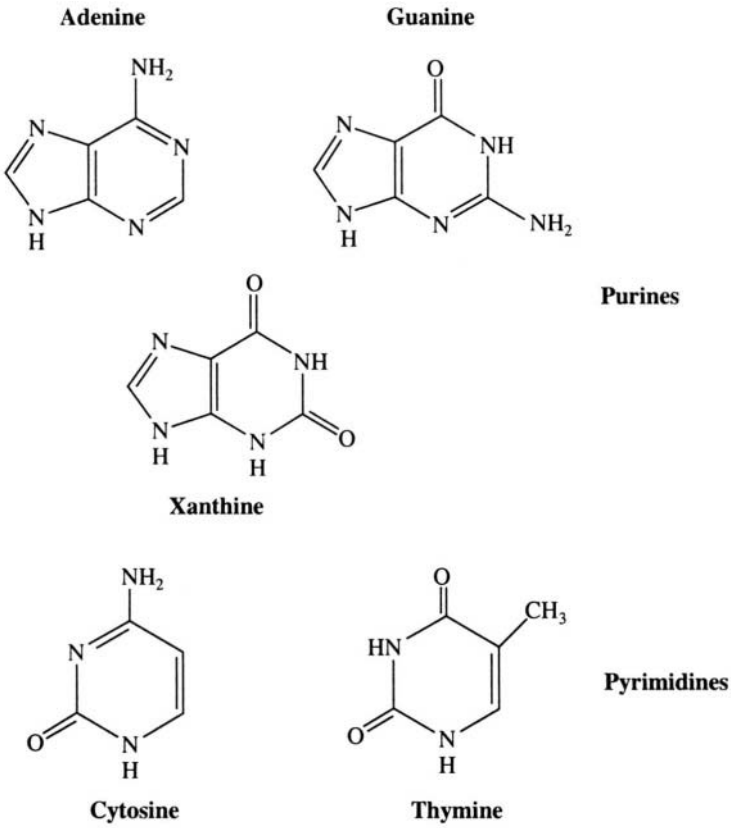


Fig. 14.4 Purines and pyrimidines

Synthesis of pyrimidines, viz. C, U and T has not been studied extensively. Action of an electric discharge on a mixture of methane and nitrogen led to the formation of cyanoacetylene. Cyanoacetylene reacts with aqueous cyanate to give cytosine in quite good yield (20%). Uracil may be obtained by hydrolysis of cytosine.

14.4.3 Sugars

Formaldehyde is one of the simplest of organic molecules that is readily formed under pre-biotic

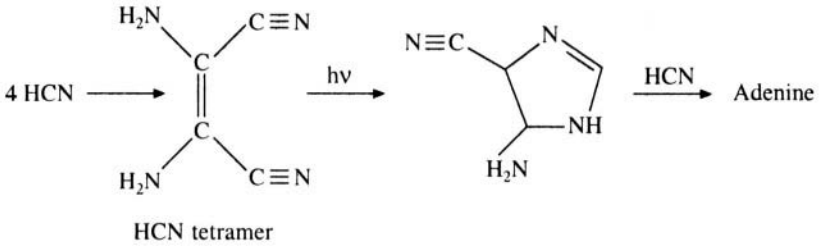


Fig. 14.5 Photochemical rearrangement

conditions, as attested by its presence in intersteller clouds. Sugars are formed by treating formaldehyde with certain alkaline catalysts like sodium hydroxide.

$$nH_2C = O \xleftrightarrow[\text{Inorganic ashes as catalysts}]{} \text{triose, tetrose, pentose, hexose}$$

While the suggested mode of synthesis is attractive, it is not without difficulties, such as the following. (a) Sugars are unstable in aqueous solution. (b) The mixtures of sugars obtained usually contain a very small percentage of ribose, which is the constituent of nucleic acids. (c) The reaction requires a minimum concentration of 0.01 M formaldehyde.

Groth and Suess also produced organic compounds by irradiating a gaseous mixture of H_2O and CO_2 during their work related to the origin of life. Similarly Garissian and coworkers bombarded an aqueous solution of carbon dioxide with some Fe^{2+} and formed formaldehyde, formic acid and succinic acid.

14.4.4 Nucleotide synthesis

Ribose exists in two different ring forms, viz. furanose and pyranose (Figure 14.6). The natural nucleosides are all furanosides, but most organic syntheses give a mixture of isomers. Further, heterocyclic bases can form bonds to a sugar at more than one position. Thus, when purines are heated along with sugars in the presence of certain inorganic salts, a mixture of nucleosides is obtained in reasonable yield. A good percentage of the yield consists of natural nucleosides. The mixture of salts obtained by evaporating seawater is one of the most effective catalysts tested so far. Two alternative approaches to the synthesis of nucleosides are possible. Either a sugar can be attached to the preformed base or a base can be constructed on the preformed sugar. The later model has been

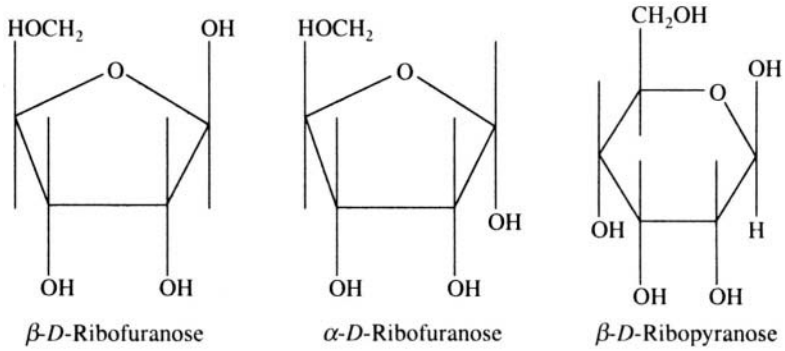


Fig. 14.6 Furanose and pyranose

accepted as the correct one in the case of prebiotic synthesis. Phosphorylation of nucleosides can be achieved by heating them with acid phospates such as NaH_2PO_4 and $Ca(H_2PO4)_2$. The reaction occurs quite rapidly at 130°C. This yields the required nucleotides. Thus, there is substantial evidence to prove that a variety of small organic molecules relevant to living systems could be synthesised under prebiotic conditions.

14.4.5 Optical activity

Most biological molecules exhibit optical activity i.e. they rotate the plane of polarisation of polarised light. In an organic synthesis reaction, optically inactive starting materials will give rise to optically inactive products. In this case, the product molecules may either be individually optically inactive or the product may contain L and D molecules in equal quantities so that the product solution is optically inactive. There is no obvious reason why molecules of one chirality should be produced more than the other in a symmetric environment. In biomolecules only L-amino acids and D-ribose (Figure 14.7) occur, even though molecules with opposite chirality have an equal chance of forming a polymer. This asymmetry needs to be explained. Among the many plausible explanations that have been offered are the following. There are several asymmetric phenomena in nature that could contribute to an asymmetric environment. For example sky light is circularly polarised to a small extent. Radioactive **β-decay** exhibits handedness or chirality. Similarly the rotation of earth, the magnetic field of earth, etc. may also contribute to the asymmetric environment and this could lead to the preference for one handedness of the synthesised molecules over the other.

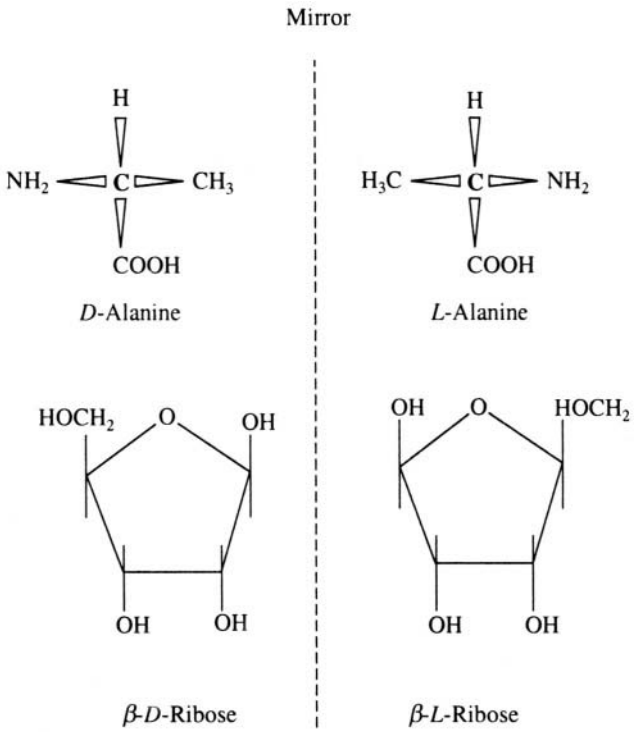


Fig. 14.7 L and D amino acids and riboses

14.4.6 Polymerisation

Once the relevant monomers have been synthesised, the second step in prebiotic chemical evolution is the aggregation or self-organisation of molecules. The following experiment demonstrates how this could have occurred for proteins. When a pure amino acid solution is heated to about 180°C, varying quantities of diketopiperazine (a cyclic dipeptide) and small amounts of polypeptides are formed. If a mixture of 20 natural amino acids is used in this experiment, all of them are incorporated in the polypeptide like materials. These materials are called "proteinoids". Most of the bonds between the different residues in proteinoids are the normal HN–CO peptide bonds. But those formed from mixtures which include aspartic acid and glutamic acid also contain α-β and α-γ amide linkage (Figure 14.8). A surprising feature is that many of the polypeptides thus synthesised begin with glutamic acid. If the proteinoids are boiled in water, a dispersion of microspheres of size 2 microns is obtained. Some microspheres are surrounded by structures that look like bilayer membranes. Some of them look much like simple bacteria under a microscope and they can be made to divide into two, or to form buds, under appropriate conditions. These proteinoids, also known as thermal polypetides, display a catalytic activity in some reactions. However, there are some facts, which argue against thermal synthesis of polypeptides as a mechanism for prebiotic polypeptide synthesis. Temperatures of the order of 150 to 180° C are required. In the prebiotic atmosphere at the time when the monomers are available, such temperature would be found only several kilometres below the surface of the earth

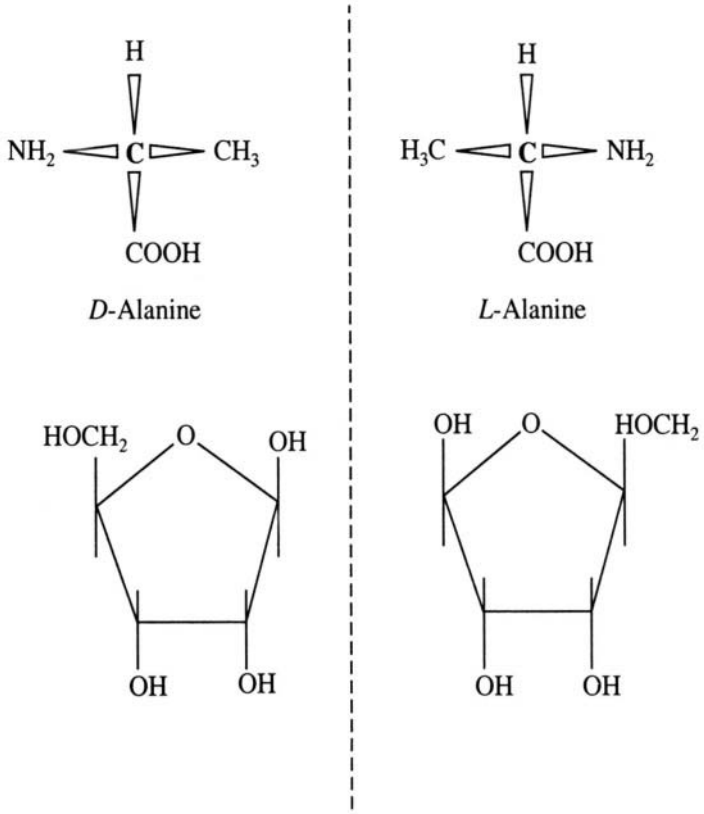


Fig. 14.8 α-β and α-γ amide linkages

or in some hot springs or in volcanoes. Other condensation mechanisms have also been proposed for condensation of amino acids into polypeptides.

The formation of short oligonucleotides from the monomers could also have occurred purely by thermal processes. However, the specificity that exists in the polymerisation of nucleotides in the present day world, viz. the $5' \rightarrow 3'$ direction of the polynucleotide chain, may not have been present under prebiotic conditions.

The prebiotic synthesis of fundamental molecules described above requires special conditions, and all of them may not have existed at the same time at the same place. For example the reactions which produce the monomers require very high concentrations of the reactants and the subsequent yield of biologically important monomers is low. It is believed that repeated evaporation and rehydration, freezing and thawing might have concentrated both the reactants and the products. The building blocks of life must have been condensed into still larger molecules by heating or chemical treatment. Thus, when the prebiotic solution of molecules (the prebiotic 'soup') cooled, it could have contained oligonucleotides, peptides and oligosaccharides. However, a mere collection of these molecules will not make a living system as we know it today.

To summarise, though the above experiments and arguments do not unequivocally establish the mechanisms for origin of life on eath, and for its subsequent evolution, they definitely show that the conditions on primitive earth were suitable for the emergence of some primitive form of life.

This Page Intentionally Left Blank

Further Reading

General

1. *Introductory Biophysics.* P. Narayanan (1999). New Age Publishing Co., Mumbai, India.
2. *Biochemistry.* D. Voet and G.J. Voet (1990). John Wiley & Sons, Inc., New York, USA.
3. *Biophysical Chemistry, Vol. I, II and III.* C.R. Cantor and P. Schimmel (1985). W.H. Freeman and Company, New York, USA.
4. *Biophysics.* W. Hoppe, W. Lohman, H. Markl and H Ziegler (eds) (1983). Springer Verlag, New York, USA.
5. *Physical Biochemistry.* D. Freifelder (1982). W.H. Freeman and Company, New York, USA.
6. *Aspects of Biophysics.* W. Hughes (1979) John Wiley and Sons, Inc., New York, USA.
7. *Biophysical Science.* E. Ackerman, L.B.M. Ellis and L.E. Williams (1979). Prentice-Hall Inc., New Jersey, USA.
8. *An Introduction to Biophysics.* C. Sybesma (1977). Academic Press, New York, USA.
9. *Molecular Biophysics.* M.V. Volkenstein (1977). Academic Press, New York, USA.
10. *Organic Spectroscopy: Principles and Applications.* Jag Mohan (2000). Narosa Publishing House, New Delhi, India.

Chapter 1

1. *Fundamentals of Physics.* D. Halliday, R. Resnick and J. Merill (1988). John Wiley & Sons, New York, USA.
2. *College Physics.* F.W. Sears, M.W. Zemansky and H.D. Young (1985). Addison Wesley Publishing Company, Massachusetts, USA.
3. *Chemical Principles.* R.E. Dickerson, H.B. Gray, M.Y. Darensbourg and D.J. Darensbourg (1984). The Benjamin Cummings Publishing Company, California, USA.
4. *Organic Chemistry.* T.L. Finar (1969). ELBS and Longmans, Green and Co., Ltd., London, UK.
5. *Stereochemistry of Carbon Compounds.* E.L. Eliel (1962). McGraw-Hill Book Company, New York, USA.

Chapter 2

1. *Principles and Techniques of Practical Biochemistry.* K. Wilson and K.H. Goulding (1986). Edward Arnold (Publishers) Ltd., London, UK.

2. *Biophysical Chemistry, Vol. II.* C.R. Cantor and P. Schimmel (1985). W.H. Freeman and Company, New York, USA.
3. *Physical Biochemistry.* K.E. van Holde (1971). Prentice-Hall Inc, New Jersey, USA.
4. *Physical Chemistry of Macromolecules.* C. Tanford (1961). John Wiley and Sons. New York, USA.

Chapter 3

1. *Principles and Techniques of Practical Biochemistry.* K. Wilson and K.H. Goulding (1986). Edward Arnold (Publishers) Ltd., London, UK.
2. *Biophysical Chemistry, Vol. II.* C.R. Cantor and P. Schimmel (1985). W.H. Freeman and Company, New York, USA.
3. *Physical Biochemistry.* D. Freifelder (1982). W.H. Freeman and Company, New York, USA.
4. *Physical Chemistry.* P.W. Atkins (1976). W.H. Freeman and Company, San Francisco, USA.
5. *Physical Chemistry of Macromolecules.* C. Tanford (1966). John Wiley & Sons, New York, USA.

Chapter 4

1. *Organic Spectroscopy.* W. Kemp (1987). ELBS/Macmillan, Basingstoke, UK.
2. *Biophysical Chemistry, Vol. II.* C.R. Cantor and P. Schimmel (1985). W.H. Freeman and Company, New York, USA.
3. *Fundamentals of Molecular Spectroscopy.* C.N. Banwell (1983). Tata McGraw-Hill Publishing Company Ltd., New Delhi, India.
4. *Molecular Spectroscopy.* G.M. Barrow (1962). McGraw-Hill Book Company Inc., New York, USA.

Chapter 5

1. *Light Microscopy in Biology.* A.J. Lacy (ed.) (1989). IRL Press at Oxford University Press, Oxford, UK.
2. *Optical Techniques in Biological Research.* D.L. Rousseau (ed.) (1984). Academic Press, New York, USA.
3. *Fundamentals of Optics.* F.A. Jenkins and H.E. White (1975). McGraw-Hill Book Company, New Delhi, India.
4. *Optics.* F.W. Sears (1964). Addison-Wesley Publishing Company Inc., Massachusetts, USA.

Chapter 6

1. *Electron Microscopy in Biology.* J.R. Harris (ed.) (1991). IRL Press at Oxford University Press, Oxford, UK.
2. *The Principles and Practice of Electron Microscopy.* I.M. Watt (1985). Cambridge University Press, Cambridge, UK.
3. *Optical Techniques in Biological Research.* D.L. Rousseau (ed.) (1984). Academic Press, New York, USA.
4. *The Electron Microscope.* M.E. Haine and V.E. Cosslett (1961). E. and F.N. Spon Ltd., London, UK.

Chapter 7

1. *Fundamentals of Crystallography. C.* Giacovazzo, H.L. Monaco, D. Viterbo, F. Scordari, G. Gilli, G. Zanotti and M. Cati (1998). International Union of Crystallography and Oxford University Press, Oxford, UK.
2. *Principles of Protein X-ray Crystallography.* J. Drenth (1994). Springer-Verlag, New York, USA.
3. *X-ray Structure Determination—A Practical Guide.* G.H. Stout and L.H. Jensen (1989). John Wiley & Sons, New York, USA.
4. *Crystals, X-rays and Proteins.* D. Sherwood (1976). Longman Group Ltd., London, UK.
5. *Crystallography and Crystal Chemistry.* F.D. Bloss (1971). Holt, Rinehart and Winston Inc., New York, USA.

Chapter 8

1. *Magnetic Resonance.* C.L. Khetrapal and G. Govil (1991). Narosa Publishing House. New Delhi, India.
2. *Nuclear Magnetic Resonance.* Atta-ur-Rahman (1986). Springer Verlag, New York, USA.
3. *Biophysical Chemistry, Vol. II.* C.R. Cantor and P. Schimmel (1985). W.H. Freeman and Company, New York, USA.
4. *Magnetic Resonance.* K.A. Mclauchlan (1972). Clarendon Press, Oxford, UK.
5. *Interpretation of NMR Spectra.* R.H. Bible (1965). Plenum Press, New York, USA.

Chapter 9

1. *Molecular Modeling.* A.R. Leach (1996). Addison-Wesley Longman Ltd. Essex, England, UK.
2. *Molecular Dynamics Simulation.* J.M. Haile (1992). John Wiley & Sons, New York, USA.
3. *Computer Graphics.* D. Hearns and M.P. Baker (1990), Prentice-Hall Inc., New York, USA.
4. *Computing for Biologists.* A. Fielding (1985). The Benjamin Cummings Publishing Company, California, USA.

Chapter 10

1. *Protein Structure.* M. Perutz (1992). W.H. Freeman and Company, New York, USA.
2. *Introduction to Protein Structure.* C. Branden and J. Tooze (1991). Garland Publishing Company, New York, USA.
3. *Nucleic Acids in Chemistry and Biology.* G.M. Blackburn and M.J. Gait (eds.) (1990). IRL Press at Oxford University Press, Oxford, UK.
4. *Biophysical Chemistry, Vol. I.* C.R. Cantor and P. Schimmel (1985), W.H. Freeman and Company, New York, USA.
5. *Principles of Protein Structure.* G. Schultz and R.H. Schirmer (1984). Springer-Verlag, New York, USA.
6. *Principles of Nucleic Acid Structure.* W. Saenger (1984). Springer-Verlag, New York, USA.

Chapter 11

1. *Biochemistry.* L. Stryer (1995). W.H. Freeman and Company, New York, USA.
2. *Biochemistry.* G. Zubay (1993). Wm. C. Brown Publishers, Iowa, USA.
3. *Biochemistry.* D. Voet and J.G. Voet (1990). John Wiley & Sons, New York, USA.
4. *An Introduction to Biophysics.* C. Sybesma (1977). Academic Press, New York, USA.

Chapter 12

1. *Biochemistry.* L. Stryer (1995). W.H. Freeman and Company, New York, USA.
2. *Biophysics.* W. Hoppe, W. Lohman, H. Markl and H. Ziegler (eds) (1983). Springer-Verlag, New York, USA.
3. *Aspects of Biophysics.* W. Hughes (1979). John Wiley & Sons, New York, USA.

Chapter 13

1. *Textbook of Medical Physiology.* A.C.Guyton and J.E. Hall (1998).W.B. Saunders Company, New York, USA.
2. *Cell and Molecular Biology, Indian Edition.* E.D.P. Robertis and E.M.F. Robertis (1995). B.I. Waverly Pvt. Ltd., New Delhi, India.
3. *Biophysical Science.* E. Ackerman, L.B.M. Ellis and L.E. Williams (1979). Prentice-Hall Inc., New Jersey, USA.
4. *Bioelectric Phenomena.* R. Plonsey and D.J. Fleming (1969). McGraw-Hill Book Company, New York, USA.

Chapter 14

1. *Molecular Biology of the Gene.* J.D. Watson, N.H. Hoplans, J.W. Roberts, J.A. Steitz and A.M. Weines (1987). the Benjamin Cummings Publishing Company, California, USA.
2. *Seven Clues to the Origin of Life.* A.G. Cairns-Smith (1985). Cambridge University Press, Cambridge, UK.
3. *Exobiology.* C. Ponnamperuma (ed.) (1982). North-Holland Publishing Company, Amsterdam, Holland.
4. *The Origins of Life on Earth.* S.L. Miller and L.E. Orgel (1974). Prentice-Hall Inc. New Jersey, USA.

This Page Intentionally Left Blank

Index

α-helix, 166
β-sheets, 166
β-strands, 166
β-turns, 167
^{13}C NMR, 122

Abbe, 82
absorbance, 35, 40, 44, 59, 68
acetylcholine, 213
actin, 200
adsorption, 24
adsorption chromatography, 27
affinity chromatography, 30
agarose, 29
alamethicin, 194
amide I, 70
amide II band, 70
amino acids, 13, 27, 60, 73, 126, 161, 169, 188, 235
Anfinsen, 161
angular momentum quantum number, 4
anomalous scattering technique, 108
antenna molecules, 179
antibodies, 31, 84, 152, 231
anti-bonding orbitals, 6
anti-Stokes lines, 72
aorta, 206
asymmetric carbon, 12
atomic force microscope, 86, 90, 93
atomic number, 3, 13, 20, 32, 55, 114
ATP, 175, 176
auricle, 206, 227
autoradiography, 23
Avagadro's number, 38, 52
axons, 208

base pairs, 11, 67, 151
batch methods, 102
beam splitter, 60
Beer-Lambert law, 59
Bekesey, 230
birefringence, 37
black body, 2
blood circulatory system, 205
blood pressure, 205
Bohr, 3, 74
Boltzmann, 16, 40, 115
bonding orbitals, 6
Bragg, 95
Bravais lattices, 96
bright field microscopy, 82
Brownian motion, 48

C_2'-endo, 145
C_3'-endo, 145
Cahn-Ingold-Prelog, 13
Calorie, 15
canonically conjugate variables, 3
capsid, 172
carbon fixation, 187
carbon-monoxide stretching frequencies, 70
cardiac muscle, 199, 205
CD, 61, 67, 134
celite, 28
cellulose, 27
central nervous system, 210, 224
centrifugal force, 39, 41

chemical energy, 174, 179, 199, 218
chemical evolution, 232, 240
chemical fixation, 90
chemical receptor, 217
chemical shift, 75, 117, 123
chirality, 11, 239
chlorophyll, 180
chloroplast, 180
Chou, 133
chromatic aberration, 87
chromatin, 154
chromatography, 24, 27
chromophores, 60, 61, 221
circular dichroism, 59, 61
circular polarisation, 59
cis-trans isomers, 11
cochlea, 227, 230
coenzymes, 178
collagen, 167
column chromatography, 26
complementary base-pairing scheme, 151
computer graphics, 95, 110, 131
concentration gradient, 40, 44, 196
conductance, 216, 223
cones, 220
configuration, 6, 90
configurational isomers, 11, 139
conformation, 11
conformational isomers, 11
conformational marker, 70
connective tissue, 202
constrained-restrained refinement, 109
continuous flow electrophoresis, 35
continuous wave, 117, 126
contrast, 56, 77, 87, 129
COSY, 127
Cotton effect, 64
counter current chromatography, 27
counter ions, 32, 38
coupled reactions, 176, 192
covalent bonds, 5, 90, 121
CPK, 137
Crick, 131, 151
crystal, 9, 76, 90, 119, 134, 154
crystal systems, 97
crystallin, 218
cytoplasm, 195
cytosol, 187

dark field microscopy, 82
Darwinian evolution, 232
De Broglie, 2
definition of life, 232
dendrites, 210
deoxyribose, 145
depolarisation, 214
detector, 51, 67, 89, 105
detergent, 34
deuterium, 20, 56, 114, 124
diaphragm, 82, 218
diastereoisomerism, 12
diastolic pressure, 205
diatomic molecule, 6, 67
dichroic effects, 70
diffraction, 2, 58, 103
diffusion, 37, 213
dipole moment, 10, 52, 67
direct methods, 108
disk of confusion, 87
distances geometry, 128
distribution coefficient, 24
disulphide bond, 169
DNA Polymorphism, 151
DNA supercoiling, 154
domain structure, 169
double helix, 61, 92, 126, 151
double resonance, 122

Ebeling, 235
ECG, 208
Eigen, 234
eigenfunction, 2
eigenvalue, 2
Einstein, 2
Einstein coefficient, 65
Einthoven's Law, 209
electric birefringence, 49, 50
electric dipole, 10
electrocardiogram, 208
electromagnetic spectrum, 58, 67
electron configuration, 9
electron density, 6, 55, 92, 107
electron diffraction, 86, 92, 173
electron optics, 86
electron spin resonance, 37, 74
electronegativity, 10, 121, 213
electrophoresis, 24, 32
ellipsoids, 39
elliptically polarised, 49, 63
ellipticity, 63
eluant, 25
enantiomers, 12

endothermic, 17, 19
enthalpy, 17
entropy, 16, 174
equilibrium method, 44
ESR, 74
ethidium bromide, 27, 67
eustachian tube, 228
evolution, 232
excitation, 22, 66, 117, 125, 129, 179, 204, 230
exergonic, 194
exothermic, 17
extinction coefficient, 59

FAD, 178, 188
Faraday constant, 19, 196
Fasman, 133
Feistel, 235
F_{hkl}, 106
Fick's law, 39
FID, 117, 126
fine structure, 65
first law of thermodynamics, 15
Fischer projection, 14
flow birefringence, 49
fluorescence, 48, 59, 65, 83, 181
fluorescence microscopy, 83
fluorescence spectroscopy, 59, 65
fossils, 234
Fourier analysis, 92, 108
Fourier transform, 71, 92, 106, 126
free energy, 17, 138, 174, 193
free induction decay, 117
friction coefficient, 38
frictional force, 32, 35, 38
FTNMR, 117, 120, 126
furanose ring, 147

'g' factor, 74
ganglia, 211
gas constant, 19, 42
gas liquid chromatography, 28
Geiger-Muller, 22
Gels, 29, 31
generator potential, 218
genetic damage, 23
geometrical optics, 77, 92, 218
glide plane, 98
glycolysis, 187
glycosidic bond, 147
goniometer., 105
G-quartet, 156
gramicidin, 194
ground state., 5, 21, 60
group transfer potential, 177
Guinier plot, 55

half-integral spin, 114
Hamiltonian operator, 2
Hans Krebs, 188
Hauptman, 108
heavy atom, 90, 107
Heisenberg, 2
Heisenberg's Uncertainity principle, 79
Helmholtz, 17, 230
high voltage electrophoresis, 33
Hodgkin, 217
hormones, 205, 230
HPLC, 27
Huxley, 217
hydrodynamic properties, 38
hydrogen bonds, 10, 70, 169
hydrophobic interactions, 128
hydroxyapatite, 27
hyperchromism, 61
hypercycle, 234
hypermeteropia, 224
hypochromism, 61

'i'-motif, 156
icosahedron, 172
image reconstruction, 91
interface diffusion, 102
interference, 2
internal energy, 16
interstellarcompounds, 233
intersystem crossing, 66
ion-exchange chromatography, 24, 28
ionising radiation, 130, 233
ionization, 22, 106
irreversible process, 15
iso electric focussing, 35
isometric, 204
isothermal process, 17

Karle, 108
Karplus equation, 121
kinetic energy, 1
Krebs cycle, 188
Kuhn, 234

Larmor frequency, 74, 115
lattice, 91, 95, 120, 129

lattice translations, 97
laws of thermodynamics, 13
least squares, 108
light harvesting, 179
light scattering, 37, 51
line broadening, 120, 125
linear combination of atomic orbitals, 6
line-width, 119
lipid bilayer, 194
liquid phase, 25, 236
liquid-liquid chromatography, 27
London dispersion, 10
low voltage electrophoresis, 33
luminescence, 83, 226

macromolecular assemblies, 92, 172
magnetic lenses, 88
magnetic quantum number, 5
magnetic resonance imaging, 112, 130
magnetic resonance spectroscopy, 130
magnetisation, 115
magnetogyric ratio, 114
mass, 1
mass number, 20, 114
matrix, 3, 27, 48, 127, 188
matrix mechanics, 3
mechanical energy, 199, 229
mechanics, 1
membrane transport, 194
meniscus, 44
meromyosin, 202
microelectrode, 213
microheterogeneity, 154
microtomy, 90
Miller, 235
Miller indices, 103
Miller's experiment, 235
mobile phase, 24
molar extinction coefficient, 59
MO-LCAO approximation, 6
molecular dynamics, 128, 138
molecular evolution, 232
molecular exclusion chromatography, 24, 29
molecular orbitals, 5
molecular replacement, 108
molecular spectra, 59
momentum, 2
moving boundary electrophoresis, 32
multi-component system, 18, 38
multi-dimensional NMR, 113, 125
multiple isomorphous replacement, 107
muscle action, 199
muscle contraction, 1, 174, 202
myelin, 210
myocardium, 206
myofibrils, 199
myopia, 224
myosin, 199

Na^+-K^+ pump, 195
NAD, 31, 178, 187, 222
NADP, 178, 184
neuronal response, 211
neurons, 210, 218
neurotransmitters, 231
Newman projection, 14
Newton's laws, 2
ninhydrin, 27
NMR, 58, 71, 95, 124, 147
NMR imaging, 120, 129
NOESY, 127
noradrenaline, 213
nuclear magnetic resonance, 112
nuclear magneton, 74
nuclear Overhauser effect, 118, 128
nucleosome, 154

oblate ellipsoid, 39
opsin, 221
optic nerve, 219
optical activity, 12, 59, 239
optical isomers, 12
optical rotatory dispersion, 37, 59
orbital, 4, 22, 60
ORD, 61, 67
origin of life, 238, 241
osmotic pressure, 37
ossicles, 227
oxidation, 187
ozone layer, 233

pair-production, 23
paper chromatography, 26
paramagnetic metal ions, 76
partial charge, 140
particles, 1
partitionchromatography, 27
partition coefficient, 28
Patterson synthesis, 108
Pauli's exclusion principle, 5
peptide backbone, 133
peripheral nervous system, 211

Perrin factor, 48
pH titration, 125
pharynx, 228
phase contrast microscopy, 82
phase problem, 107
phosphocreatine, 199
photon, 2
photo-neural transducer, 224
photophosphorylation, 186
photoreceptor, 219
photosynthesis, 174, 179
photosynthetic unit, 179
photosystem, 181
phototrops, 174
Planck, 2
Planck's constant, 2
point groups, 97
polarisability, 52, 72
polarising microscopy, 84
polyacrylamide, 29
polynucleotide, 61, 74, 145, 234, 241
postsynaptic fibre, 212
potential, 3, 9, 17, 33, 92, 128, 177, 192, 204, 231
prebiotic evolution, 235
prebiotic synthesis, 235
preparation of the specimen, 90
presynaptic fibre, 212
primary structure of the protein, 160
primitive earth, 235
principal quantum number, 4
prolate ellipsoid, 39
propensity, 134
prosthetic molecule, 60, 182
protein fold, 133
protein folding problem, 133
proteinoids, 240
proton gradient, 185
proton motive forced, 193
puckered, 145

quanta, 2
quantum, 2
quantum mechanics, 2
quartenary structure of the proteins, 161, 170
quasi-equivalent, 173

R factor, 109
radiation, 2
radiative lifetime, 66
radioactive decay, 20
radioactivity, 20
radioimmunoassay, 23
radioisotopes, 20, 23
radiotracer, 23
radius of gyration, 53
Raman, 71
Raman effect, 71
raster pattern, 88
Rayleigh scattering, 52, 71
receptor potential, 223
redox, 19
reducing agents, 35
relaxation parameters, 119
replication, 91
re-polarisation, 214
residual rotation, 62
resolution, 34, 79, 109, 154
resonance absorption, 58
resonance condition, 76, 113
resonance frequency, 74, 114
respiration, 176, 188
resting potential, 214
retinal, 221, 225
reverse phase, 27
reversible process, 15
R_f values, 27
rhodopsin, 221
ribose, 145, 175, 238
RNA, 27, 67, 144, 172, 234
rods, 138, 220
Rossmann, 108
rotating frame, 117
rotational friction, 48
rotational spectra, 58
rotor, 43
R-S convention, 13
Rutherford, 3

sarcolemma, 204
sarcomere, 200
sarcoplasm, 202
SAXS, 54
scanning electron microscope, 88
Scherage-Mandelkern, 48
Schlieren optical system, 44
Schrodinger, 2
Schrodinger wave equation, 3
screw axis, 98
SDS-PAGE, 34
second law of thermodynamics, 16, 174
secondary structure of the protein, 161
sedimentation, 37

sedimentation velocity method, 44
self-organisation, 234, 240
SEM, 88
semi-empirical, 65, 128, 139
sensitiser molecule, 180
sequence specific structure, 154
short contacts, 145
signal to noise ratio, 117
Simha factors, 47
skeletal muscle, 199
sliding filament model, 201
small angle x-ray scattering, 54
smooth muscle, 199
solar energy, 174
solid phase, 25
somatic receptor, 217
space groups, 97
spatial theory of hearing, 230
specific rotation, 62
specific volumes, 38
specimen preparation, 90
spectrum, 21, 60, 94, 113, 221
spherical aberration, 87, 218
spike potential, 213
spin quantum number, 4, 75, 114
spin state, 112
spin-density, 129
spin-labels, 76
spin-lattice, 119
spin-spin coupling, 120
spin-spin coupling constants, 121
spin-spin relaxation, 119
splitting factor, 74
spontaneous process, 16
spontaneously emitted radiation, 66
state functions, 16
stationary phase, 25
stereochemistry, 11, 59, 108, 154
stereoisomers, 13
steric contacts, 11
stimulated emission, 65
Stokes law, 38
Stokes lines, 72
striated muscle, 199
structure amplitude, 109
structure of NaCl, 9, 95
supersaturated state, 95
supersecondary structure, 169
Svedberg equation, 44
Swann cells, 210
symmetry, 10, 95, 172
symmetry operations, 97
synapse, 210, 223
systolic pressure, 205

T_1, 75, 129
T_2, 75, 129
teleceptor, 217
TEM, 87
tendons, 202
tensor, 48
tertiary fold, 128
tertiary structure, 161, 169, 234
tetanus, 203
tetraplex, 156
thermodynamical system, 13
thermodynamics, 13
thin layer chromatography, 24, 26
third law of thermodynamics, 16
transducer, 202, 224
transduction of sound, 227
transition, 5, 60, 66, 113, 126
transmission electron microscope, 87
triplex, 156
tRNA, 144, 158
tropomyosin, 201
troponin, 201
tunnelling electron microscope, 86,92
tympanic membrane, 227
tympanum, 227

ultracentrifuge, 41
unit cell, 96
universal principles, 1
Urey, 235

valinomycin, 194
valves, 206
van der Waals interactions, 10
van der Waals radius, 11
vapour diffusion, 102
vector cardiography, 209
ventricles, 206
vibrational levels, 59, 65
virus structure, 172
viscosity, 37, 46
vision, 1, 219
visual receptor, 218
voltage clamp, 215
voltage gated channels, 214

Watson, 131, 151

wave function, 2
wave like phenomena, 1
wave packets, 2
wave-particle duality, 2
weak interactions, 10

x-ray crystallography, 95
x-ray scattering, 37, 54

Yphantis, 45

Zimm Plot, 53
zonal sedimentation, 45
zone electrophoresis, 33